Engaging Social Media Users:

Understanding Online Interactions in a Retail Context

Jamie Mahoney

PhD

December 2019

This is dedicated to all those - family and friends - who have walked
this journey with me, for however long, and helped me to
get to where I am today.
To those who are no longer here, we miss you.
This is for you.

Acknowledgements

First, I would like to thank my supervisors, initially at the University of Lincoln, and then at Northumbria University – Professor Shaun Lawson, Dr Duncan Rowland, and Professor David Kirk for their advice and support throughout my studies.

I would also like to thank friends and colleagues at both the Lincoln Social Computing Research Centre (LiSC), and Northumbria Social Computing Lab (NorSC) for their friendship, support, and guidance over the years. To those that kept me sane – and caffeinated – through the years (you know who you are), thank you.

Thanks must also go to Sister London, for their partnership throughout this process, and for providing the industry expertise in guiding this programme of research.

Finally, I must thank my family, who have all supported me throughout this process. A huge thank you to my wife, Rachel, whose continued support and encouragement has kept me motivated throughout, and without whom I could not have got through these past few years.

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted, initially by the Ethics Committee at University of Lincoln in 2013, with further ethical approval sought and granted by Northumbria University in January 2017.

I declare that the word count of this thesis is 60,494 words.

Abstract

As existing social media platforms are further developed, and new platforms are introduced, their use by both organisations and individuals continues to grow. The adoption of such platforms as a means for commercial (e.g., retail) organisations to interact with clients and customers is now ubiquitous. There remains an ongoing need however to understand how and why organisations and individuals continue to interact through social media in the ways that they do. Prior literature has highlighted the need to understand not only what social media behaviours should be implemented by organisations, but how user engagement behaviours persist, develop, and differ across multiple social media platforms. Based on these identified gaps in the literature, the primary over-arching aim of the research described in this dissertation is to develop a greater understanding of social media users' engagement behaviours and the effect that this engagement may have on maintaining and growing online social media audiences.

This thesis presents a sequence of four studies that utilise both qualitative and quantitative approaches to understand social media users' engagement behaviours in real world contexts. All studies took an empirical approach to social media data collection and analysis, taking into account the rapidly changing ethical and privacy landscape that has developed during the course of the work. The first of these studies offers an outline of how retail organisations make use of social media, the types of content that they share, and the resulting levels of engagement they receive. This knowledge is then further developed in later studies, demonstrating and evaluating a means by which user engagement with different types of content can be modelled. Further to this, a longitudinal study demonstrates how networks of account followers develop over time, and how this growth may be modelled and predicted, including levels of engagement as one such potential predictor. Though primarily focused on these academic contributions, the dissertation also provides a number of practical implications for organisations. These include insights into approaches to develop a greater understanding of their online audiences, and the ability to tailor their online content strategies by understanding how groups of followers engage with different types of content. The thesis concludes with implications for future work, highlighting how recent literature and the findings of these studies can be used to motivate continued research in this field.

Contents

List of Tables

List of Figures

Glossary of Terms and Definitions

Betweenness Centrality
Betweenness centrality "*measures the extent to which a vertex lies on paths between other vertices*." [143, p.185]

Bridge
"*An edge that joins two nodes ... in a graph is called a bridge if deleting the edge would cause [the two nodes] to lie in two different components. In other words, the edge is literally the only route between its endpoints*." [60, p.46]

Centrality
Centrality relates to how 'involved' a node is within a network [200]. Many centrality measures exist; see: Closeness Centrality, Betweenness Centrality, Transitivity and Eigenvector Centrality for specific examples.

Coding
"In qualitative data analysis, the assignment of simple words or short phrases to capture the meaning of a larger portion of (the original) textual or visual data. Whether or not supported by computer software, the analyst must make the coding decisions for every item, including what to code and how." [213, p.334]

Closeness Centrality
Closeness centrality "*measures the mean distance from a vertex to other vertices*." [143, p.181]

Community
Communities, within a network, are groups of vertices that are much more tightly connected within the community than they are to nodes outside of the community.

Cutpoint
"*A node, n_i, is a cutpoint if the number of components in the graph that contains n_i is fewer than the number of components in the subgraph that results from deleting n_i from the graph. ... In a communications network, an actor who is a cutpoint is critical, in the sense that if the actor is removed from the network, the remaining network has two subsets of actors, between who no communication can travel*." [200, p.112-113]

Degree
"*The degree of a vertex in a graph is the number of edges connected to it*." [143, p.133]

Diffusion
Diffusion models and processes are often used "*as a simple model of spread across a network, such*

as the spread of an idea or the spread of a disease." [143, p.152]

Discourse Analysis
Discourse analysis is a research method for understanding spoken or written content, particularly in relation to its social context. In doing so, it aims to understand how language is used in in real-life situations.

Edges
Edges are the connections between nodes or vertices in a network. See also: Links.

Eigenvector Centrality
Eigenvector centrality is a method of determining the centrality of a user within a network. This approach is summarised in that "*...not all neighbours are equivalent. In many circumstances a vertex's importance in a network is increased by having connections to other vertices that are themselves important. This is the concept behind eigenvector centrality. Instead of awarding vertices just one point for each neighbour, eigenvector centrality gives each vertex a score proportional to the sum of the scores of its neighbours*." [143, p.169]

Goodness of Fit Measures
These measures, which include Deviance, Akaike's Information Criteria (AIC) [156] and Nagelkerke's R^2, indicate how well the model that has been created fits the entire dataset that was used to generate it. These measures are useful when comparing one model to another. The deviance metrics indicate the goodness of fit of the model, comparing the 'null' model with one including the variable being tested (hence two values). For the AIC value, a lower value indicates a better fit, while a higher Nagelkere's R^2 indicates a better fit.

Graph
"*A graph is a way of specifying relationships among a collection of items. ... Graphs are useful because they serve as mathematical models of network structures.*"[60, p.21]

Grounded Theory
Grounded theory is a research method primarily concerned with the generation of theory 'grounded' in data that has been collected and analysed systematically.

Homophily
"*A tendency within a network for nodes to be disproportionately tied to others who share one or more salient statuses with them.*" [50, p.15]

Indegree
The indegree of a vertex in a graph is the number of edges connecting into the vertex.

Interpretative Phenomenological Analysis
IPA is a qualitative research method which aims to provide detailed understandings of personal, lived, experiences.

Links
Links are the connections between nodes or vertices in a network. See also: Edges.

Mixed Methods Research
"Deliberately designing a study to use quantitative and qualitative methods, both of which are needed to address the research questions of interest."[213, p.337]

Network
In its simplest form, a network is "*a pattern of interconnections between a set of things*." [60, p.2]

Network Density
"*The number of ties in a network expressed as a proportion of the total number that are possible*."[50, p.15]

Network Partitioning
In its simplest form, network partitioning relates to the division of a network into smaller groups.

Nodes
The 'actors' within a network - the objects, people, or 'things' that have demonstrable links or relationships between them. See also: Vertices/Vertexes.

Non-parametric tests
In comparison to parametric tests, non-parametric tests do not make assumptions about the distribution of data. Non-parametric tests are often utilised when the assumptions of parametric tests are violated.

Outdegree
The outdegree of a vertex in a graph is the number of edges connecting from the vertex.

Qualitative Research
Qualitative research involves the collection and analysis of non-numerical data, such as text, video, and audio.

Parametric tests
Parametric tests are statistical tests which make assumptions about the distribution of the data. As such, the tests are statistically valid, under these conditions.

Quantitative Research
Quantitative research involves the collection and analysis of numerical data, for statistical analysis.

Selection
"*The tendency of people to form friendships with others who are like them is often termed selection, in that people are selecting friends with similar characteristics*." [60, p.81]

Self-Categorisation Theory
"*Self-categorisation theory focuses on how people categorise themselves and other (including the definition of social identities) and how this categorisation process serves as a basis for group behaviour*." [72, p.6]

SIR Model
A model of diffusion where actors can be in one of three states: (S) susceptible to being 'infected', (I) infected, or (R) recovered.

SIRS Model
Similar to the SIR model, but with a fourth stage included - after going through the (R) recovered stage, actors return to being (S) susceptible to 're-infection'.

Socialisation
Synonymous with 'social influence', suggests that "*existing links in the network service to shape people's (mutable) characteristics*." [60, p.82]

Social Influence
Social influence suggests that "*existing links in the network service to shape people's (mutable) characteristics*." [60, p.82]

Social Network
"*A social network consists of a finite set of actors and the relation or relations defined on them. The presence of relational information is a critical and defining feature of a social network*." [200, p.20]

Sociometry
A field of research, attributed to Moreno, that has since developed into what is now known as social network analysis.

Effect Size
The effect size, or estimated effect size, when used in logistic regression models, indicates the magnitude and direction of the relationship between the independent and dependent variables. This is presented in a 'raw' form, in the same unit of measurement as the independent variable, and as such the numbers are not necessarily directly comparable between cases where the variables are measured in different units, or have different upper or lower limits.

Transitivity
"*The idea of transitivity suggests that any two nodes are more likely to enjoy a tie if each is tied to a common, third party*." [50, p.15]

Vertices / Vertexes
The 'actors' within a network - the objects, people, or 'things' that have demonstrable links or relationships between them. See also: Nodes.

Chapter 1

Introduction and Overview

1.1 Background and Motivations

The use and prevalence of social media continues to increase, with estimates in early 2019 placing the number of social media users at 3.484 billion worldwide (a 9% increase from 2018), with approximately 4.388 billion individuals accessing the Internet [201]; organisations in almost every sector continue to develop their use of social media to help achieve their goals – be that in politics [45, 77], social and health support networks [132], education [80, 96], or retail and branding [112, 120, 162]. With this increasing use comes increasing costs, both temporal and financial [172], as social media use, campaigns, and strategies are further integrated into organisations' existing operating processes.

Academic research into the use of digital media, including social media, in the field of retail and branding is well established, with many noted benefits. One example being the possibility for social media content to cross geographic boundaries often experienced with traditional media, reaching a much wider audience [174], ultimately resulting in increased brand recognition [66], sales performance [106], and longer-term brand loyalty [129]. While the integration of technology has not necessarily changed the underlying goals of marketing communication [172], the means by which these goals can be achieved continue to develop. This changing landscape creates a need to understand how individuals engage with such media, and to develop approaches that facilitate this understanding, and that can be adapted to different contexts in which organisations may operate.

Research focusing more specifically on social media use often considers what online engagement may look like, and underlying causes as to why individuals choose to engage with online content at all [26]. Mechanisms for engagement differ between social media platforms, with individuals often maintaining a presence on multiple platforms. With traditional marketing approaches often focusing on market segmentation as a means of identifying varying needs and motivations, there is also much research which considers determining the characteristics or demographics of social media users [158, 159, 165, 176, 177], and how these may impact their use of social media [165]. Techniques and approaches, taken from the field of social network analysis (SNA) are often integrated into these studies, using concepts such as homophily to explain behaviours, identifying groups of highly connected individuals [19, 144], and using network positioning to identify 'key' individuals within a network

[113].

Combined, these existing areas of research, and the continued growth of social media use raises questions such as *"how can organisations make more effective use of social media, create more effective social media content strategies that grow their online audiences and benefit the organisation, helping it to meet its goals?"*.

The main motivation for this thesis, therefore, is to facilitate the creation and implementation of improved social media strategies through developing a greater understanding of how social media users interact and engage with online content and how online audiences' growth can be modelled and predicted. As a result of this, in a practical context, strategies can be developed and refined to respond to the changing preferences of online audiences, enabling organisations to make the most of the opportunities provided by online social media. With existing research linking positive online presence to organisational performance [55, 130], the ability to tailor online strategies to a known and profiled audience, thus improving performance, will be of benefit to organisations.

It would be impractical to consider, in detail, social media use across multiple sectors within this work, and as such, and further motivated by ongoing collaboration with an industry partner, a single sector, that of retail organisations, has been selected. To further narrow the focus, geographically-bound retail locations have been selected to act as case studies. These locations, such as shopping streets and shopping malls or centres are often owned by a single organisation, with their online and social media presence maintained by a single team, either '*in-house*' and thus directly controlled, or contracted out to a third-party organisation. These locations present an interesting and complex use case for investigating social media use, as their social media presence has to represent not only the interests of the retail location as a whole but should also represent the business interests of their tenant retailers. These tenant retailers will often operate across a range of differing product markets and have differing target demographics. As such, the social media audience for these retail locations is likely to be diverse and will continue to change and develop as new tenant retailers join the retail location, and other tenant retailers leave. Such diversity suggests that the online audience for each retail location may have different characteristics and developing an understanding of these will be beneficial in the development and implementation of social media content strategies. It also suggests that any approaches or techniques will have to be flexible or adaptable to be applicable in multiple contexts.

Although the studies presented here have a specific focus in terms of the types of social media account being considered and analysed, the presented findings, and documented techniques can be utilised within many other contexts and should not be considered as only appropriate for use solely within the context presented here.

1.2 Aims and Objectives

The aim of this research was to develop a greater understanding of social media users' engagement behaviours, and the effect that this engagement may have on maintaining and growing online social media audiences.

In order to achieve this aim, the following research objectives have been identified:

1. To develop an understanding of how retail locations make use of social media, including the range of content being shared and how this may affect the levels of engagement from their social media audiences.

2. To develop a greater understanding of how online social media audiences may grow (or decline) and develop over time.

3. To investigate the extent to which social network analysis techniques and metrics can be used to indicate the likely organic growth or decline of social media audiences.

4. To implement and evaluate a method of generating profiles of social media users' engagement with specific content, to enable retail locations to understand the aggregate behaviour of their online audiences.

5. To develop an understanding of the role of engagement with social media content in the growth and development of social media audiences.

Each of these objectives, which themselves are motivated through a combination of work with an industry partner and identifying 'gaps' in prior research are addressed within this thesis, with some being addressed within a single chapter, while others are addressed throughout. A summary of the related research, the identified 'gaps', and how these relate to these objectives is provided in Section 3.5. In addressing these objectives, multiple research contributions are made, which are outlined in the next section of this chapter. These will also have potential implications for practice, which are discussed in more detail in Chapter Nine and Chapter Ten.

1.3 Overview of Contributions

This thesis presents a means of developing a greater understanding of the engagement preferences of social media audiences, and how the potential future growth (or decline) of these audiences can be better understood and predicted. While each chapter that presents a study will state the contributions made as a result of that study, the overarching contributions of the thesis are presented here.

First, and foremost, a technique for developing user engagement profiles is presented and evaluated. Such profiles can be grouped, demonstrating groups of users that engage with social media content in similar ways; these groupings can also be used in conjunction with, and to supplement, other social network metrics. These profiles, and groupings of users, demonstrate to social media account managers, for example, the types of content that are likely to elicit the greatest engagement from their audiences, or the content that is likely to encourage engagement from those accounts that themselves have the largest audiences, or those seen to be influential or key in some other way. As outlined previously, this knowledge will allow social media content strategies to be refined and developed, so that engaging content for a known audience is produced and shared, leading to a more effective use of social media, which is likely to ultimately benefit the organisation.

Second, a greater understanding of how social media audiences develop over time is presented. Discussed initially in Chapter 5, and then further in Chapter 6, the presented findings demonstrate that different audiences grow at different rates, and in different ways. Further, while ‘traditional’ social network metrics can be used, to some extent, as indicators to predict this development over time, it is suggested that individual preferences for consuming and engaging with social media content may also play a role in the longitudinal development of social media audiences.

Finally, the contribution of engagement behaviours to the growth and development of online audiences is presented. Building on the findings presented in Chapter 6 and utilising the method presented in Chapter 7, engagement behaviours of individual accounts and those around them in the follower network are evaluated as a means of inferring when and if users are likely to ‘leave’ the network – i.e. to stop following a particular account. Again, if this knowledge is applied in a practical context, organisations may develop their understanding of how these engagement behaviours, and the previously discussed social network metrics, may indicate when a user may be starting to leave the network. This would then allow, for instance, content strategies to be refined (particularly if the users are of high value to the organisation) to keep these users engaged and, ultimately contribute - albeit indirectly - to both the online and offline performance and success of the organisation.

1.4 Structure of this Thesis

The remainder of this thesis is structured as follows:

- In order to provide a broad context and background for the literature that follows in later chapters, Chapter 2 explores social network analysis, providing an overview of the development of this field, as well as key concepts and theories, such as community detection, social influence and information diffusion models.

- Chapter 3 provides an overview of academic research relating to the behaviour of organisations, and their use of technology for marketing, including social media, building on some of the theories and concepts outlined in the previous chapter. This literature is used to inform and contextualise the studies that follow; this chapter concludes with a section that relates the results of this literature review to the research objectives previously outlined in this chapter, in Section 1.2.

- Chapter 4 details the methodological approaches used within the studies presented within this thesis, including data collection and analysis techniques, as well as relevant ethical and legal considerations, examples of ‘best practice’, and the ethical approval processes undertaken during the studies presented in this thesis.

- Chapter 5 details an initial study to understand how retail locations utilise social media, in terms of the content they produce, how this behaviour may develop over time, and the resultant levels of engagement from their audiences. This study demonstrates: that different retail locations utilise social media in different ways, which may indicate conscious decisions in terms of developing and implementing social media strategies; that audiences engage in different ways and to different levels, and that changes and developments within the audience itself may contribute towards differing engagement behaviours from the audience as a whole. Part of the work presented in this chapter was published in [127].

- Chapter 6, building on the findings of the study detailed in Chapter 5, considers how social media audiences may develop over time - sometimes referred to as follower churn. Based on social graph data collected over a 16-month period, this study investigates the use of 'traditional' social network analysis techniques to model or predict the potential growth (or decline) of online audiences. This study demonstrates that some metrics can be used, to some extent, to indicate individual audience accounts that are at risk of unfollowing the account, or are susceptible to following the account in the near future. It also suggests that consideration of other factors, including how individual users engage with particular content, may supplement more traditional social network analysis techniques, providing a more robust means of modelling and predicting social media audience growth.

- Chapter 7 focuses on the development, implementation and evaluation of a technique for profiling user engagement, on a large scale, with the social media content produced and shared by organisations' social media accounts. A method for clustering these profiles together is suggested, with the resultant groups of similar users (based on engagement behaviours) demonstrated, and compared with groupings of users produced using more traditional social network analysis techniques. This study demonstrates that grouping users that have similar engagement behaviours is possible (in a similar vein to more traditional market segmentation techniques), and produces different groupings to other analysis techniques, which has implications for relying solely on traditional, long-standing SNA techniques when developing social media content strategies.

- Chapter 8 builds on the studies detailed in both Chapter 6 and Chapter 7, by considering engagement behaviour as an indicator for predicting and modelling social media audience growth. This study demonstrates that engagement, when considered in isolation, is not effective in predicting follower churn. When used as supplementary predictors in other models however, the efficacy of these models is improved. Such findings highlight the need for continued research in the area of understanding the motivations for actively engaging online.

- Chapter 9 draws together the findings of the studies presented in Chapters 5, 6, 7 and 8 and outlines the contributions of each study individually, and the body of work as a whole. With these contributions outlined, the chapter proceeds to outline how such techniques and findings might be utilised by organisations, both within the sector used as an example within this thesis, and within other sectors. This chapter also presents a short summary of relevant literature that has been published during and after the work presented in this thesis was conducted.

- Chapter 10 draws the thesis to a conclusion, through a review of the aims and objectives, reiterating where these have been addressed in the presented studies, and a summary of contributions to knowledge and practice.

1.5 Summary

This chapter has provided an introductory framing to this thesis. After an overview of the background and motivations for this work, the aims and objectives have been outlined. Following this, the contributions offered by the research have been summarised, as well as the structure of this thesis. In the next chapter, relevant work surrounding social network analysis is discussed, providing a background to the literature discussion that follows in Chapter 3, as well as the studies presented in later chapters.

Chapter 2

Network Theory and Analysis

In this chapter, various aspects of social network analysis are discussed. This chapter, therefore, describes some of the techniques, concepts, and metrics that will be drawn upon in the studies presented within this thesis, providing the relevant background to these studies. This chapter is outlined as follows: first, an historical overview is provided, demonstrating the development of social network analysis as an area of research. Following this, key theories and concepts are explored in more detail. A glossary of key terms and their definitions can be found towards the beginning of this thesis.

2.1 Early Foundational Work and Development into a Recognised Discipline

Social network analysis (SNA) is not a new concept, but its perceived importance has re-emerged with the seemingly global adoption of social media platforms in businesses and organisations, as well as by individuals. In this section, the historical roots of SNA are briefly outlined, before modern theories and approaches, and their applicability to the studies presented here, are discussed in more detail.

Though not labelled as such, the roots of SNA can be traced back to the late 19th and early 20th centuries [59, 175]. This early work focused predominantly on the structural perspective in the study of human behaviours [211].

Durkheim, for example, discussed what is now referred to as social influence (see Glossary of Terms and Definitions), noting that:

> *It is generally accepted today, however, that most of our ideas and our tendencies are not developed by ourselves but come to us from without.*[59, p.22]

This work laid the early foundations for this area of research, and the underlying principles and focuses, such as how social structure plays a role in behaviour, can be seen in modern approaches and theories. Credited with laying the foundations for what we now know as social network analysis is Jacob Moreno [136, 137, 138]. Along with Helen Jennings [139], Moreno developed a technique known as sociometry (see Glossary of Terms and Definitions), which was defined as:

> *the mathematical study of psychological properties of populations, the experimental technique of and the results obtained by application of quantitative methods.* [138, p.15-16]

The underlying basis to sociometry is that individuals make choices in terms of their relationships with others, and that these choices have consequences that may or not be foreseen by the individuals involved. Moreno again describes this:

> *Choices are fundamental facts in all ongoing human relations, choices of people and choices of things. It is immaterial whether the motivations are known to the chooser or not; it is immaterial whether [the choices] are inarticulate or highly expressive, whether rational or irrational. They do not require any special justification as long as they are spontaneous and true to the self of the chooser. They are facts of the first existential order.* [138, p.720]

Discussed in more detail later, we can see the relevance of these statements by both Durkheim and Moreno to the study of online networks, such as social media platforms. Individuals on these platforms make decisions, perhaps without any conscious thought, as to who to follow, what content to engage with, and which pages to like. These decisions then impact what content, information, and opinions they are exposed to. This content can then feedback into their future decisions, which again impacts which accounts they decide to follow, and what content they engage with. Written decades ago, these concepts still appear to hold true, despite being conceived and written in a time when online networks were far from even being thought of.

After the development of sociometry, this field of work experienced what is referred to by many, including Freeman [70], as a 'dark age'. During this time, it ceased to be a focus of the social sciences, and could not be identified as either an approach to data collection and analysis, or indeed a distinct theoretical perspective.

However, during this time, work on social network analysis, in one form or another, continued 'quietly' [211] at several universities in the United States of America. Much of this work was conducted in labs that were either led by, or involved, many individuals whose names are now easily recognisable in this field.

Focus on social network analysis re-emerged in the 1970s when scholars from Harvard University, including Harrison White, published work on structural equivalence and 'blockmodelling' [204]. Amongst others, White trained individuals including: Edward Laumann, Peter Bearman, Ronald Breiger, Kathleen Carley, Bonnie Erickson, Claud Fischer, Mark Granovetter, and Barry Wellman. As well as being responsible for important developments in sociology and related fields, they also took up positions at other universities, which may have contributed to the more widespread focus on this field of research.

Social network analysis, which started off as a multidisciplinary effort [107, 143], has benefited from these roots and continues to cross the boundaries of many disciplines. The scope of social network analysis now also includes fields such as political science, economics, psychology and sociology [211]. While the application of these concepts may differ depending on the context, the same underlying theories and concepts are applicable. Some of these theories and concepts, which are applicable within the studies presented in this thesis, are discussed in the following subsections.

Figure 2.1: A demonstration of a simple network, consisting of nodes and edges.

2.2 Fundamentals of Network Theory

Network theory is the underlying theory and principles related to the use of graphs (networks) as a representation of the relationships between objects. In this context, networks consist of two main elements - nodes (or vertexes), and edges (or links). Here, nodes represent objects, with edges used to represent some form of relationship between them. Figure 2.1 demonstrates a basic example of nodes connected by edges.

Network theory has been applied and used in many contexts. These include, but are not limited to, physics, biology, economics, engineering, and sociology. In each of these contexts, the application of the theory and principles will change, but the underlying theory and principles remains consistent.

There are many terms that are used in relation to network theory; a list of relevant terms and their definitions can be found in the glossary towards the beginning of this thesis. Key aspects of of this theory and its application are discussed in the following sections.

2.2.1 Centrality

Often one of the crucial uses or motivations for the application of network theory is the identification of key or important individuals, or nodes, within the network. This apparent need is not new, as the predecessors to terms used in modern network-related research can be found in the works of Moreno [136], where the notions of sociometric ‘stars’ and ‘isolates’ are discussed.

The context of this importance will often vary depending on the dataset from which the network is formed, or the situation which the network represents. However, mathematically, one of the ways in which the importance of a node within a network is determined is through the use of *centrality* measures, of which there are several.

First developed in the 1940s and 1950s by individuals such as Bavelas [15, 16], the notion of centrality relates to the ‘involvedness’ of individual nodes within a network. Rather than focus on whether the node is heavily involved in either receiving or sending information (identified with metrics such as indegree and outdegree), centrality focuses on their overall involvement within the network. Definitions of specific methods of determining centrality are outlined in the glossary towards the beginning of this thesis.

This concept of centrality can play an important part in identifying key nodes in organisations’ follower networks. However, Wasserman and Faust do raise an important point when considering exactly what such measures actually demonstrate, and the usefulness of doing so.

But the major question still remains unanswered: Are the nodes in the graph centre ... the most "central" nodes in a substantive sense - that is, does the centre, or centroid, of a graph contain the most important actors? In part, this is a question about the validity of the measures of centrality - do they really capture what we substantively mean by 'importance' or 'prominence'? Can we simply focus on the actors who are 'chosen' the most to find the most important actors? [200, p.170]

Within the scope of this thesis, this point is both complex and important - there may be scenarios and contexts in which the centrality of a user may not be crucial to the ongoing growth and success of organisations' social media networks, nor their social media advertising strategies.

For example, if the goal of an organisation is numerical growth of their online social media following, is a central figure within the existing network of particular importance for them? It could be argued that their importance in the growth of the network could also depend on their popularity outside of the existing network, and as such, further account information would have to be taken into consideration. If the most central users only have followers within the network, and very few 'other' followers which they could encourage to join the network, then their overall importance could indeed be lower than a somewhat less central account that has a far greater number of 'out-network' followers, that they could encourage to join the network.

In contrast to this, if the organisation has different goals, and perhaps intends to identify those that could be key contributors to conversation and information flow within the network, then centrality may be of importance to them. Encouraging discussion from those that are heavily involved within the existing network could be key. It could also be argued, however, that other metrics may be of more importance in this context - indegree for example (i.e. nodes that have high numbers of followers within the network) could be a more useful indication of their use in the context of encouraging information distribution through the network.

Situations such as those outlined above suggest that organisations need to be clear on their goals, particularly when deciding on which metrics to use. This would be of particular importance if key advertising and marketing strategy decisions were made based on the results of, in this case, social network analysis techniques.

2.2.2 Network Partitioning and Community Detection

Social networks, such as those constructed to represent the relationship between followers of a particular social media account, can be, and often are, partitioned into smaller networks. These smaller networks - communities - often represent groups of accounts or individuals that are more strongly or tightly connected to each other than those outside of the group.

The partitioning of a network in these contexts is determined mathematically. Methods such as those proposed by Girvan and Newman [78], function based on the premise of removing *key edges* from the network. These edges are identified as such due to carrying the highest 'traffic' when looking at the shortest paths between each set of nodes. As such, they become vital for connecting various regions of the network [60]. By removing these edges, the network can be partitioned into separate communities.

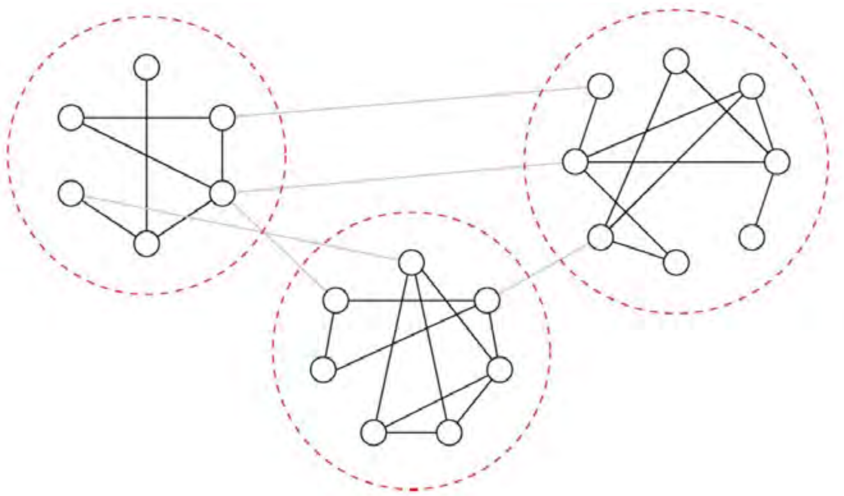


Figure 2.2: A simple demonstration of community detection within a network. In this example there are three communities, identified within the dashed circles, which have dense internal links, but between which there is a lower density of external links. Image and description based on Newman and Girvan (2004) [144].

A simple example of this process is demonstrated in Figure 2.2 (taken from [144]). In this example, the three identified communities (indicated by the red dotted boundaries) can be seen clearly. Within each of the identified communities, the density of links (or edges) between the nodes is much higher than the links between them.

It is often suggested that communities detected in this manner demonstrate the concept of *homophily* - that is, that individuals within a network tend to connect more to those that are similar to themselves [200] (see Glossary of Terms and Definitions). As a result, the groups of individuals in these communities are similar in some manner - be that demographics, behaviour, geography etc. As stated previously, many of the concepts within social network analysis are by no means new; this notion of homophily can be traced as far back as the writing of Plato *"similarity begets friendship"* and Aristotle when he writes that people *"love those who are like themselves"* [60, 131].

Two main processes related to homophily are of particular interest here: that of *selection* - who we chose to form relationships with, and *socialisation* (or *social influence*) - how and why individuals may (or may not) alter their behaviour to be similar to that of those around them [71, 103] (see Glossary of Terms and Definitions).

There are numerous studies that have applied this concept of homophily across a range of disciplines. Farver, for example, considered the role of social relationships (and homophily) in aggressive behaviour in young children [65], while Himelboim, McCreery and Smith [88] considered networks and homophily in the exposure to various ideological and political content in online networks. In both of these studies, it was found that network structure and played an important role. Farver found that similar behaviour was observed within groups of children, and made suggestion for targeting groups, rather than individual children, in order to attempt to improve behaviour. In a similar vein, Himelboim, McCreery and Smith [88] found that networks of connected users were often exposed to the same kind of political and ideological viewpoints, with little *"cross-ideological interaction"* between these groups of users.

While there are many examples of the use of these concepts within research, what is not always evident is whether individuals change their behaviour due to the social networks they find themselves in, or individuals with similar behaviour seek each other out in order to form the observed relationships. Easley and Kleinberg make an interesting observation.

> *Such questions [on the interplay between selection and social influence] can be addressed by using longitudinal studies of a social network, in which both the social connections and the behaviours within a group are tracked over a period of time.* [60, p.82]

Easley and Kleinberg elaborate further on the role of homophily.

> *The point is that an observation of homophily is often not an endpoint in itself but rather the starting point for deeper questions - questions that address why the homophily is present, how its underlying mechanisms will affect the further evolution of the network, and how these mechanisms interact with possible outside attempts to influence the behaviour of people in this network.* [60, p.83]

This opportunity afforded by longitudinal studies, and the fact that merely identifying homophily within a network is not an end goal, are utilised within the studies presented within this thesis in order to understand how behaviour and network growth may influence one another in online social networks.

2.2.3 Bridges and Cutpoints

Within the field of network theory, Wasserman and Faust define a bridge:

> *A bridge is a line that is critical to the connectedness of a graph. A bridge is a line such that the graph containing the line has fewer components than the subgraph that is obtained after the line is removed. ... The removal of a bridge leaves more components than when the bridge is included.* [200, p.114]

Easley and Kleinberg provide a plainer description of a bridge:

> *An edge that joins two nodes ... in a graph is called a bridge if deleting the edge would cause [the two nodes] to lie in two different components. In other words, this edge is literally the only route between its endpoints.* [60, p.46]

In essence, bridges act as the only edge (or line) between two sub-graphs, or communities, within the network. If the bridge is removed (i.e. one user stops following another), then the means of communicating between the two sub-graphs is also removed. To paraphrase [60], in relation to the simple network in Figure 2.3, while the nodes A,B,C and D may tend to be exposed to similar opinions and similar sources of information, B's link to E may provide the only link to other information,

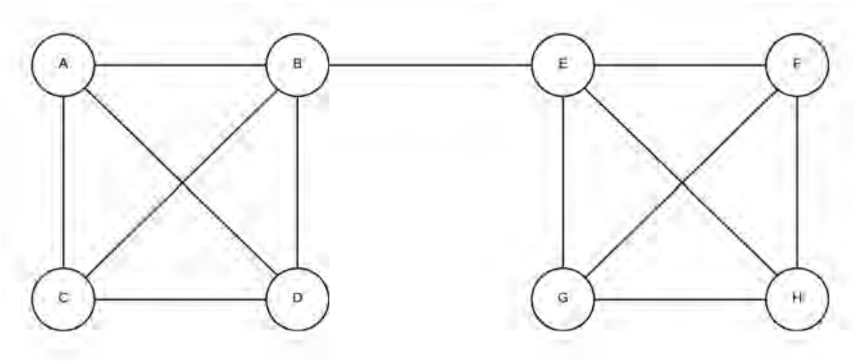


Figure 2.3: A demonstration of bridges and cutpoints in a simple network. If the edge between nodes B and E was removed, the graph would be partitioned, and any direct information flow between the two would be removed. Likewise, if node B or E was removed from the network, then the same effect would take place. Adapted from [60].

opinions and sources that they wouldn't otherwise hear about. This again highlights the potential importance of such relationships between pairs of nodes.

The type of relationship explained previously is often referred to in the phrase "*the strength of weak ties*" . Whilst groups may be tightly connected, they can often be fairly insular, and the 'weak' tie - in this case, the tie between node B and E - can be particularly advantageous, as it not only links the two components together, but can also be the only way for information to flow between two otherwise entirely separate groups of nodes. In such a situation, the strength does not necessarily lie in the groups of heavily connected individuals, but rather in the single relationship tie between two nodes.

The concept of 'cutpoints' is heavily related to that of bridges:

> *A node, n_i, is a cutpoint if the number of components in the graph that contains n_i is fewer than the number of components in the subgraph that results from deleting n_i from the graph. ... In a communications network, an actor who is a cutpoint is critical, in the sense that if the actor is removed from the network, the remaining network has two subsets of actors, between who no communication can travel* [200, p.112-113].

Bridges and cutpoints (also defined in the Glossary of Terms and Definitions) are demonstrated in a simple example graph in Figure 2.3. Once communities of connected users are identified, the identification of bridges and cutpoints becomes an important action. Defined in simple terms, cutpoints and bridges are nodes and edges that are the means by which communities within a network are linked. Removal of one of these would mean that the two communities of users are no longer linked.

Identifying potential bridges and cutpoints can be crucial in some scenarios. For example, in information flow contexts, losing a crucial node or edge can prevent information from flowing from one sub-graph to another. When dealing with the spread of a contagion, treating or inoculating a node either side of a bridge (i.e. person, community, etc) can help prevent the spread from one community to another.

Within the context explored in this thesis - that of promotional content on online social media - identifying bridges and cutpoints (i.e. individuals, or their social media accounts) can be crucial in successful online marketing - being able to identify those that may disseminate information to a wider audience, or another network, can be key.

2.3 Social Influence and Group Behaviour

Social influence can play an important part in the success of online social networks, particularly when employed for organisations' gain - whatever that 'gain' may be. When used in terms of marketing and advertising, organisations may wish to make use of social media and social networks to influence individuals to behave in a particular manner. This may take the form of brand recognition, influencing present or future purchasing decisions, or encouraging long-term brand loyalty. Such influence may stem directly from the organisation's own social media presence, or by encouraging positive behaviour from those within their networks to influence those around them.

It is not just enough to identify central or influential individuals, how 'influenceable' those around them are must also be taken into consideration.

> *However, a characteristic feature of interpersonal influence is that actors' attitudes and opinions may reflect the preferences of others who occupy different social positions than their own (hence, children may reflect the attitudes of their parents and ancestors, adolescents may reflect the attitudes of friends and friends-of-friends who have a different socioeconomic status than their own, and workers may reflect the attitudes of supervisors and owners).* [60, p.513]

Concepts such as this suggest that those in different positions within the network - perhaps those loosely connected, on the fringes of the network - may take cues from more central, highly connected, or otherwise 'successful' accounts within the network.

One particular theory that is related to this, and may play a role in influencing the behaviour of individuals, or groups of individuals, is 'self-categorisation theory'. This concept is defined by Friedkin & Johnsen (though not originally developed by them):

> *Self-categorisation theory focuses on how people categorise themselves and others (including the definition of social identities) and how this categorisation process serves as a basis for group behaviour.* [73, p.6]

This theory, therefore, could be seen as developing the notion of homophily somewhat, in that it is no longer similar individuals that group together, but rather those that perceive themselves as being similar. This perception of similarity may then influence how an individual chooses to behave.

> *Persons define themselves as belonging to a particular social category, ascertain the norms of that social category, and bring their attitudes and behaviours into conformity with those norms.* [73, p.7]

Based on this perception of the self and the group, prototypes of the 'typical' group member are often formed. As Hogg writes:

> *Prototypes ... are embodied as a reified image of a 'most prototypical' group member - an ideal or representative category member. ... The prototype is the cognitive representation of the group norm or the group stereotype.* [91, p.94]

Such prototypes may be more clearly defined in situations where the group is formed around a strong concept, opinion, or ideal, and could be found more readily in contexts such as activism, political parties, or religious groups. What remains to be seen is whether such prototypes are evident in contexts where the groups are more loosely formed. Hogg goes on to discuss the idea of these prototypes being shared between group members:

> *Since prototypes are, by definition, shared among group members, one consequence of the depersonalisation process is relative intra-group uniformity of perceptions, attitudes and behaviour. In this way the self-categorisation process accounts for conformity to group norms.* [91, p.94]

As mentioned previously, in situations where the 'group' is not formed around a concrete concept or ideal, then the notion of a 'group' existing at all (in the minds of the group members) may be unlikely. If so, this would impact the likelihood of a group prototype existing, or being shared between members, and as such there is unlikely to be a strong prototypical behaviour within the group.

> *If there was not a shared perception of the distribution of persons' attitudinal or behavioural positions, then there might not be a shared definition of the group prototype.* [73, p.8]

It could be strongly argued that for a prototypical behaviour to exist within a group, that the group members must be consciously aware that they belong to that group. However, definitions used and provided by scholars such as Friedkin & Johnsen do not require that group members perceive themselves as belonging to a group. Rather, they build on definitions provided in earlier works, defining groups as a "*collection of individuals who are interacting with one another*" (see also [21, 69, 92, 183]), as well as a "*collection of individuals who are interdependent*" (see also [47, 116]) that also influence each other (see also [173]) [73, p.17].

Regardless of whether an individual consciously considers themselves to be part of a particular group or groups, the individuals to whom they are connected in some way will play an important part in what they may be exposed to, such as particular points of view or information. As such the structure of the groups around an individual are likely to play an important role in both what they are exposed to, and how this may influence their future behaviours.

> *In order to understand the social dynamics of particular groups, we must grapple with the actual, realised network structures of these groups.* [73, p.11]

2.4 Diffusion Models

Diffusion models are often used to help explain or demonstrate a range of phenomena, from the spread of contagious diseases, through to participation in social movements [185], and are now being applied (to varying levels of success) to behaviour on online social media platforms. Many of these models took the form of, or are explained in the context of, contagion and epidemiology - measuring, determining, and modelling the spread of epidemics - such as the SIR or SIRS models [208, 216]. When applied to social networks and social media, these models suggest that users are susceptible (S), infected (I), or recovered (R) from a particular action (with users returning to a susceptible state in the SIRS model).

Applied to the spread of information online, such models would suggest that all users are susceptible to sharing a piece of information. Once a particular activation event has occurred - such as seeing a particular person sharing that information, or finding the information useful or amusing - the user may be 'infected', and share that information. The user would then 'recover' and potentially become 'susceptible' again, depending on the model being used. While this may provide a somewhat simplistic, if still useful, model for the spread of information, it does not perhaps take into account many nuances and features of online social networks that may contribute to the spread of information online.

Other approaches take into account concepts such as influence (described earlier), as well as positioning within a group, and network structure [58]. These concepts, when applied to earlier models, could act as contributing factors to what 'activates' the user to progress from the susceptible stage to the 'infected' stage. Regardless of the model used, or referred to, Easley and Kleinberg make a crucial point:

> *One of the fundamental things we learn from studying diffusion is that there is a crucial difference between learning about a new idea and actually deciding to adopt it.* [60, p.509]

Relating this to the context of this thesis, an individual or group of individuals simply learning about a retailer, retail location, or particular product line may often not be deemed a success by that organisation. There is still a need to act on that exposure to new information, such as making a purchase, for example. If we solely consider individuals' online behaviours, and the notion of diffusion of information, then it is not merely enough to expose social media users to information by posting something online, through Facebook or Twitter for example. Organisations should seek to take advantage of the features and affordances of these platforms, encouraging users to share the information more widely, and act upon it, in order for the organisations to reach a wider audience and ultimately benefit the organisation further. Part of this process will be to understand what 'activates' people within their own networks (i.e. online followers) in order to encourage them to engage and share information further.

2.5 Summary

Literature discussed in this chapter has included some of the underlying concepts and theories related to social network analysis, as well as broader concepts such as social influence and models of diffusion within the context of network structures.

This has demonstrated that social network analysis is not a new concept and has deep-seated roots through various research areas and traditions, dating back many years. Though it is now often - and seemingly 'easily' - applied to social media data (perhaps due to the abundance and variety of data currently being made available to researchers and practitioners), the underlying concepts and theories are often originally based on offline interactions between individuals. What remains unclear is whether such techniques and concepts can be readily and accurately applied to online interactions in all contexts, and whether these can be used in modelling and predicting the growth of online communities that do not necessarily find themselves bound by the temporal, geographical, and inter-personal constraints that would be applicable in the more 'traditional', offline, scenarios in which these theories and concepts were originally formed.

In the next chapter, organisational behaviour in terms of the adoption and use of technology, including social media, is discussed. Some of the concepts explored within this chapter - such as information diffusion - are discussed within this context. In doing so, the motivations and contributions of this thesis are further contextualised.

Chapter 3

Organisational Behaviour and Use of Digital Technology for Marketing

In this chapter, research relating to the behaviour of organisations - particularly the use of technology - is outlined and discussed, drawing on multiple relevant domains. In doing so, this thesis is further contextualised within existing research and current practices. As outlined previously, the work presented in this thesis is situated within the domain of retail organisations; as such, while existing research will be discussed and presented from other domains, its applicability to the retail domain will be highlighted where appropriate. This chapter concludes with a summary of identified motivations and 'gaps' in the explored research, and how these relate to the research objectives of this thesis.

3.1 Organisations' Need to Understand their Market and Audience

Organisations, across all sectors, have a need to understand their audience, or target market. This need is not new but the potential size of an organisation's audience has been increased with the ongoing prevalence of technology [174].

Though not a universally adopted approach, in many cases this understanding takes the form of, and is a result of, 'market segmentation'. Emerging as an area of both research and organisational practice in the 1950s, market segmentation continues to be one of the most researched topics within marketing literature [202].

As stated by Smith:

> *Market segmentation involves viewing a heterogeneous market (one characterised by divergent demand) as a number of smaller homogeneous markets, in response to differing preferences, attributable to the desires of consumers for more satisfaction of their varying wants.* [180, p.6]

This is further developed, more recently, by Palmer :

> *Most companies face markets which are becoming increasingly fragmented in terms of the needs which customers seek to satisfy. Segmentation then, is essentially about identi-*

> *fying groups of buyers within a market-place who have needs which are distinctive in the way that they deviate from the 'average' customer.* [155, p. 60]

This segmentation of the market can take, and has taken, many forms and has been based on many different data sources. The exact form and method of segmentation will of course differ from market to market, and will depend greatly on the context in which the organisation or organisations are operating. Regardless of the differences in how the segmentation is arrived at, there are six general principles of successful segmentation that have been put forward multiple times [12, 68, 109, 125], (Palmer [155] refines these to four, which are almost directly related to the following six and which still hold true). These principles demonstrate criteria by which the resulting segments of the market can be judged as being effective and profitable: *identifiability*, *substantiality*, *accessibility*, *stability*, *responsiveness,* and *actionability*. Each of these will now be outlined in turn.

Identifiability. This criterion relates to the extent to which distinct groups within the market can be recognised, using criteria deemed to be appropriate. This criterion is related to the third question posed by Palmer [155, p.64]: "*Can the market segment be measured?"*. Regardless of how such a criteria is phrased, this is key, in that individuals and the segment overall should be identifiable and distinct from other segments. Knowing the size and composition of each segment will allow segments to be compared, and their profit potentials (however these may be defined and determined) to be assessed [155].

Substantiality. Segments will meet this criteria if they are substantial enough to represent a large enough portion of the market to be of use. For example, a segmentation base that results in a segment of only a few people is unlikely to be of any use to the organisation. This criterion is similar to one suggested by Palmer [155, p.63], who posed the question: *"Are the segments of an economic size?"*. Developing this point further - "*Any basis for segmentation should yield segments which are of a size that a company can profitably exploit*"[155, p.63].

Palmer goes on to point out the potential issue with defining useful or effective segment sizes - segmentation of a market is essentially a balancing act. Smaller segments are likely to more accurately model the desires, attitudes or characteristics of those within the segment, but larger segments are potentially easier to target.

> *Companies face a dilemma here, because as segments get smaller, they get closer to achieving the marketing philosophy of satisfying each customer's needs as though they were the centre of the all the company's attention. The problem for the company is that smaller segments may be uneconomic to provide for.* [155, p.63]

This will be an important point to take into account when considering methods of understanding an online audience - while developing very small, targeted segments may allow for users to be more accurately modelled, the implications for implementing actions based on this may be too costly.

Accessibility. Accessible segments are those which can be reached by marketers, or other elements of the organisation, through "*promotional or distributional efforts*" [202].

This criterion is also repeated by Palmer [155], with the question "*Are the segments accessible to the company?*". Expanding this point further:

> *There is little point in going to a lot of effort in defining segments of a market where those segments are not accessible to the company, nor ever likely to be.* [155, p.64]

There are many reasons that a market segment may be inaccessible, although many of these are reduced significantly, or removed entirely, when considering online social media. There are some contexts where segments of a market may be inaccessible due to legal prohibitions or restrictions - such as the inability to trade in particular countries; buyers in a given market may be tied to particular suppliers by long-term contracts; although possible, targeting a particular segment may have cost implications that are prohibitive.

Responsiveness. This criterion is met if the identified segments respond to any marketing efforts that are targeted at them. This criterion is particularly important in determining any segmentation efforts - segmentation is likely to be costly in terms of time and finance, segments that do not respond to any resultant efforts are of no use to an organisation.

Stability. It follows that identified segments need to be stable. If a market is segmented, then the segments need to be stable enough, and for long enough, for organisations to take appropriate action. If these segments are ephemeral, then organisations will be acting on out of date, incorrect information and the effectiveness of this approach is likely to be significantly reduced. Others, such as Hooley et al. expand on this point further, highlighting that stable or predictable segments are easier to predict and assign value to, thus making them easier to assign longer-term value to [93].

Actionable. This final criterion states that any identified segment needs to be actionable - that is, organisations must be able to act on the information provided through this segmentation. If an identified segment provides no useful information in terms of business objectives, then it is of little use to the organisation.

Though readily applied to for-profit organisations, market segmentation and a need to segment and understand a market, or audience, is not limited to for-profit organisations. Institutions such as museums are also looking to develop a greater understanding of their 'market' or 'audience' as a means of adapting in an ever-changing business environment where such institutions are often at risk of closure. In some cases, this understanding goes little further than segmenting an audience into 'goers' and 'non-goers' [166]. However, the general principle behind this segmentation, or other methods used to understand a market remains the same - the need to find out more about attracting new, and retaining existing, audiences [166]. This process of market segmentation can form part of a "*post-modern marketing focus*" [166] which not only includes market segmentation, but understanding a market's (or audience's) needs, wants, attitudes and behaviours, as well as changes in these over time.

3.2 'Lurkers' and Contributors

One other means of segmenting a market, audience, or community is based on how, or if, individuals contribute to that community. 'Lurkers' and 'lurking' have been given many different definitions, with researchers sometimes providing their own criteria for identifying lurkers within a particular context or study [61, 186, 207]. The general consensus between all of these varying definitions, however, is that lurkers are responsible for very few, if any, active or demonstrable contributions to the community in which they are positioned. These individuals, regardless of the specifics of the definition being used, make up the majority of a community's members [61, 186]. In the context of social media, this may take the form of following accounts and reading their content, but not posting or responding to that content in a visible manner.

Many studies have looked upon this kind of behaviour in a negative light [108, 141, 203], although others have noted that this is not necessarily a bad behaviour that needs to be changed [61]. Indeed, some studies have shown that lurkers are not '*selfish free-riders*' [149, 150, 186, 205]. Yang, Liu and Juang (2017) note that "*lurkers should not be regarded as passive members in an online community. They are still actively evaluating the perceived community support despite being silent*"[212].

In what is often referred to as 'the 1% rule', or the '90-9-1 rule', Nielsen [146] posits that 90% of users within a community are 'lurkers', with little or no contribution, 9% of users contribute sporadically, with the remaining 1% responsible for the vast majority of contributions to the network or community. This behaviour of individuals is not limited to online social network platforms, however, with other studies also considering the participation inequality summarised in the 90-9-1 rule in contexts such open-source projects [76], and online health support networks [195].

The presence of lurkers in a community does not pose a threat to the longevity of that community, particularly in large and active online communities [212]. Rather than treating lurkers as non-users [61], these individuals should be seen as a vital part of the community [207], allowing the online content to diffuse to a wide audience [212]. "*They not only enlarge the size of a group as an audience, but they also increase its influence, as the information gained may be used in other online groups or offline settings, lead to connections with other networks, as well as bring new contacts and members*" [61].

Lurkers and contributors may join online communities for similar reasons, but they behave in different ways once they have joined the community, to fulfil their differing needs [186, 212]. Lurking provides an important way to join a community [150], and may be sufficient for users to fulfil their needs – searching for and gathering information, for example [186].

This inequality between users' participation levels is not new, and is not limited to a single online social network, or platform. As noted by Nielsen, "*the first step to dealing with participation inequality is to recognise that it will always be with us. It's existed in every online community and multi-user service that has ever been studied*" [146]. Understanding the context of the community is important for organisations, and determining whether the value of the community will be increased by turning lurkers into posters [61]. If so, then this indicates that organisations need to develop a greater understanding of the factors that drive and encourage online participation in the relevant communities [186]. These factors are likely to differ from community to community, as they are based on the individuals that make up that community. These are "*intrinsic incentives that drive people to participate in an online community and these factors are closely related to the personalities of users and their purpose of joining the community*" [186].

Pointing to the need for further research in this area, Sun, Reu, Ma (2014) state: "*Moreover, the effectiveness of motivational strategies should be further evaluated, and suggestions are needed for methods to select suitable strategies to motivate contributors according to the nature of the online* *community*" [186].

3.3 Understanding and Predicting Market 'Churn'

As outlined previously, many organisations recognise the need to strike a balance between retaining existing customers, and gaining new ones [110, 166]. This changing nature of a market or audience is often referred to as 'churn'. Organisations operating in many sectors will have a desire to understand, or even predict, market churn. In this context, churn relates to individuals that have a *"high propensity to attrite"* [196]. Although this may take on different forms depending on the market the organisation operates in, it essentially relates to identifying those that are at risk of no longer being a 'customer'. This may take the form of individuals who cancel their contracts or subscriptions, move to a competitor, or perhaps a marked drop in engagement or activity with the brand or organisation.

The cost of acquiring new customers has been shown to be much higher than the cost of attempting to retain existing customers [153, 167],with some estimates putting the cost between five [110] and twelve [190] times higher. This financial incentive, and other obvious business benefits of retaining customers [168] has led to customer churn prediction and modelling being seen as a vital part of business intelligence and operational processes [11, 53]. The act of attracting new customers is not enough to encourage organisational success, existing customers also need to be retained, as summarised by Kotler and Keller [110], adding new customers to an organisation that constantly loses a lot of customers is like *"adding water to a leaking bucket"*. This focus on customer churn is not limited to the commercial sector, however, with academics also considering churn purely as a modelling and prediction problem.

Existing research into customer churn prediction and modelling spans many business sectors, including (but not limited to) mobile telecommunications [53, 153, 167], the banking sector [79, 115, 209], insurance companies [82, 83], and Internet service providers [104]. In related work, the need to balance an organisation's activities between retaining existing customers and acquiring new ones is highlighted within the museum sector [166]. As outlined previously, the precise definition of churn will differ from sector to sector, between organisations within the same sector, or even different departments within the same organisation [164].

In each of these examples, the primary goal of customer churn prediction is to identify those at risk of 'churning', in order to put business retention procedures into place. The nature of these retention methods will of course differ between sectors, and indeed organisations. The commonalities between these approaches would be to contact those identified or put special offers and incentives in place to attempt to retain them as customers [167, 196]. In some situations, the number of identified 'churners' may be high, with differing potential costs and benefits for retaining each individual. As such, it may be necessary for organisations to make a determination as to which of their customers are worth attempting to retain, based on a variety of data sources that they may have at their disposal.

This notion of market churn links back to market segmentation approaches discussed previously. Understanding and modelling the process of churn within an organisation's target audience will facilitate a greater understanding of how stable each of the identified segments are, and perhaps indicate the point at which the segmentation process needs to be repeated in order to remain effective.

3.4 Use of Information Technology in Organisational Marketing

In the previous section of this chapter, literature and practice relating to understanding an organisation's audience or target market has been outlined. In this section, the use of technology within organisations, and particularly their marketing approaches, is discussed.

3.4.1 Adoption of (or Failure to Adopt) New Technology

The adoption, or proposed adoption, of new technologies into established organisations often faces some form of resistance [46]. Adoption of new technologies is often seen by those in managerial positions as a means of acquiring a competitive edge within a market, be that through offering new or different products or services, or by reducing costs through replacing or augmenting the workforce. In recent years, major technological changes and developments have occurred at a seemingly rapid pace. With this, there are numerous examples of major organisations that failed to adapt to, or adopt, these technologies and have failed as a result.

Kodak, the camera company, for example, filed for bankruptcy in 2012 [111]. There are two commonly accepted contributing factors to this bankruptcy. The first was the failure to maintain a prominent position in the market in the late 1980s, when roll-film was still the dominant technology. This was further exacerbated by the failure or reluctance to move into the digital market early enough - a reluctance partly due to the unknown impact this would have on their roll-film business. Ultimately, this failure to adapt to a changing market led to the eventual bankruptcy filing.

Another example of a failure to change an organisation's strategy in light of a changing market is that of Blockbuster, which filed for bankruptcy in 2010 [81]. In 2000, Blockbuster maintained a physical-store model, with patrons being charged per unit to rent movies (either VHS or DVD), with late fees being imposed for tardy returns to the store. Meanwhile, a startup company - Netflix - was beginning to operate as an online-only service, where customers (who paid a monthly subscription) could hire movies which would be sent to their home address. Late fees did not form part of their business model, customers merely had to return the film by post in order to borrow another. Despite being given the opportunity to work collaboratively in 2000, Blockbuster thought the Netflix model "too niche", and rejected the offer. The need to adapt was eventually recognised by the CEO, who attempted to change the business model in 2004 - 2005. This proposed change encountered significant resistance from within the organisation, with the proposed changes being "too costly to profitability". The CEO was eventually forced from his position, with the new CEO cancelling the proposed changes. Five years later, the organisation filed for bankruptcy.

These are just two examples of the impact of failing to adapt to, or adopt, new technologies. While there are examples of technologies that have failed, and organisations that have benefited from resisting a speedy adoption of these technologies, these highlight the need for organisations to carefully consider the impact of failing to adapt within their own markets, or move into new, emergent markets.

This is not just limited to specific markets, but also applies to technologies that have a much broader use, and (potentially) impact in many contexts and markets. Writing in 2000, Palmer questioned the growing adoption of the Internet, and the sustainability of online retail.

> *Sceptics have argued that many of the hopes held out for the Internet are over-exaggerated. If use of the Internet increases significantly, it will eventually become cluttered and the advantages gained by pioneer uses of the medium diluted. For many types of goods, customers may place great value on being able to inspect goods physically prior to purchase, something which will never be completely possibly with 'virtual' representation of goods through electronic media. The very act of visiting stores brings significant social benefits to many personal buyers. Finally, as home delivery of many goods and services (such as bread and milk) declines due to rising costs, doubts have been raised about the economics of armchair shopping in which goods ordered through a home terminal are delivered to the buyer's home.* [155, p.535]

This again highlights that it is not always (or at all) possible to accurately forecast the exact impact of a particular technology, or indeed its longevity. As such, perhaps the focus should be to remain flexible - in organisational processes - wherever possible, in order to adapt to an ever-changing technological landscape. As such, developing approaches that can be adapted to new platforms, technologies, or contexts is of increasing importance to organisations operating on digital platforms.

3.4.2 Adoption and Development of Online Presence

Since the commercial use of the Internet became widespread in the mid-1990s, organisations have sought to establish and develop their online presence to communicate information, and ultimately benefit the goals of their organisation [18]. The advantages of this technology to organisations were recognised quickly, the adoption of an online presence was seen to: potentially lower the cost of distribution (particularly for digital content), empower the consumer, and enable easier capture of customer (or potential customer) information [90].

The interactive nature of web technologies was recognised as being conducive to word-of-mouth marketing [89, 90], an area now recognised in its own right - electronic word of mouth marketing (eWoM) [101]. The adoption and use of 'early' web technologies, such as email, forums, and message boards by members of the public facilitated the indirect electronic word of mouth marketing relating to organisations. The importance of concepts such as eWoM was further emphasised by the introduction and widespread adoption of social media.

3.4.3 Adoption of Social Media

Writing in 1999, Darcy DiNucci [56], is credited as coining the term 'Web 2.0', envisioning a future where the use of the Internet developed from an easily recognisable use of websites and web browsers, to one with no "*visible characteristics at all*". Expanding this point further, DiNucci explained:

> *On the front end, the Web will fragment into countless permutations with different looks, behaviours, uses, and hardware hosts. The Web will be understood not as screenfuls of text and graphics but as a transport mechanism, the ether through which interactivity happens.* [56, p.32]

Less than a decade later, in 2005, Tim O'Reilly reflected on this notion of 'Web 2.0':

> *Web 2.0 is the network as platform, spanning all connected devices; Web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects through an "architecture of participation," and going beyond the aged metaphor of Web 1.0 to deliver rich user experiences.* [152]

Social media is one of the more recognisable tenets of Web 2.0, and has experienced its own growth and development in recent years. Various platforms are now available, each of different sizes, offering a different range of features. As social media platforms began to become more mainstream, and attract critical numbers of individuals, organisations began to recognise the need to maintain an active presence on these platforms, where they deemed it appropriate.

For organisations, social media can provide multiple benefits in order to help them achieve their goals. Social media platforms can act as both a means through which to disseminate their corporate message and related information, as well as a vital information source for gauging public response and opinion, as Zeng et al. summarise:

> *For-profit businesses are tapping into social media as both a rich source of information and a business execution platform for product design and innovation, consumer and stakeholder relations management, and marketing. For them, social media is an essential component of the next-generation business intelligence platform.* [214, p.14]

It could now be argued, however, that this is not merely limited to 'for profit' organisations, but also for non-profit and charitable organisations and communities. The pervasive nature of social media platforms make them appear to be ideal advertising and communication platforms for organisations.

Where maintaining and developing a presence on social media platforms may have started as a curiosity, or peripheral function of the organisation, many markets and organisations are now recognising the necessity of actively maintaining and developing their online social media presence as part of their main marketing processes. This point is developed further by many researchers and practitioners, who reiterate that simply having an online presence is no longer adequate. With social media platforms enabling and encouraging real-time discussion between individuals, organisations can often find themselves without control as to how they are discussed online.

> *...it is getting harder for these institutions to stay in control. Therefore, the best way for them is to be a part of social media, which is where the target audience resides.* [17, p.18]

Many well-known, large organisations were early adopters of social media, integrating it into their business practices at an early stage, and developing and refining their strategies over time. Examples of this include both Dell and Dominoes Pizza, both of which successfully adopted a social media presence prior to 2010 [87].

There are many ways in which organisations may use public-facing social media [169] (as opposed to social media platforms which are intended solely for use within the organisation). Such reasons include:

- Gauging marketplace reactions to their own services or products, or those of competitors within the wider market [102].
- Engaging (either directly or indirectly) with user communities [119].
- Managing customer relationships [134].
- Marketing products or services [193].
- Selling products and services through 'social commerce' [182].
- Collaborating and 'co-creating' with customers and consumers [84, 198].

As outlined previously, social media adoption by organisations continues to rise, with organisations either seeking to gain a competitive advantage in doing so, or doing so in reaction to their competitors online presence. Despite varying markets and underlying goals the main aim of adopting such an online presence is to spread their marketing message, whatever that may be, to a wider audience, and as such increase the standing and performance of the organisation. One of the ways in which this can be achieved is through 'word of mouth' (WoM) or 'electronic word of mouth' (eWoM) marketing.

3.4.4 (Electronic) Word of Mouth Marketing

One practical application of social influence and information diffusion theories and models (discussed in Chapter 2) is the concept of 'word of mouth' (WoM) marketing, which relates to the spreading of information from one individual to another, and has roots within 'traditional' marketing, before being expanded to 'electronic word of mouth' (eWoM) marketing to consider the spread of information in an electronic or online context.

Harris provides a simple definition of word of mouth marketing:

> *It is when one person talks to another person he or she knows about a product, service, event or issue with which one or the other has had some experience.* [86, p.10]

This definition demonstrates how word of mouth marketing and information flow through a network - be that offline or online - are based on the same concepts. One could argue that, when applied to online social networks, that the 'knowing' aspect of Harris' definition must be much more relaxed. Whilst this definition, and the concept as a whole, is rather simple, Harris elaborates further to draw attention to the dilemma that can arise due to the apparent simplicity of the approach.

The very simplicity of word of mouth advertising creates a dilemma

- *Because it is so inexpensive to promulgate, it is often discounted as insignificant.*
- *Because it is so basic to human activity, conversation is often forgotten as a medium of promotion susceptible to direction.*
- *But because it can work so effortlessly, it takes some delicacy to make it function properly.* [86, p.11]

Whilst traditionally the spread of information through WoM was limited by deliberate communication between individuals, or small groups, the emergence of online networks has removed these constraints. As such, eWoM is no longer bound by physical proximity, geographic and cultural boundaries, or temporal constraints. Online networks facilitate the sharing of messages, information, and opinions to large audiences, often in an asynchronous manner and on a one-to-many 'broadcasting' basis.

Though often discussed within the context of social media, eWoM ultimately relies on the underlying structure of networks for the successful dissemination of information and opinion. The three points from Harris, outlined previously, are clearly applicable to online social media. Though readily available to all - both individuals and organisations - the simplicity of the underlying mechanisms of these platforms can make the adoption of such platforms seem simple and straightforward, though the reality may be far from this.

3.4.5 Understanding Motivations for Social Media Use

As well as developing their own online presence, organisations need to grasp the motivations for their audiences' use of social media. The rate of social media use by individuals appears to continue to grow, with the introduction and development of new and existing platforms. There are many reasons that individuals may make use of social media platforms, such as Facebook, Twitter, and Instagram. These disparate motivations and observed behaviours have been noted by various scholars in a range of contexts.

Naaman et al. [142], for example, observed that users can be often divided into two groups, with distinct behaviours - those that use social media to talk and share about themselves, and those that relay information from other sources. This is reinforced by other studies [215], which have found that occasional Twitter users make use of the platform as a news source. Users in such a case are often referred to as data consumers, rather than producers. This, of course, may have implications when considering how (and if indeed it is at all possible) to encourage social media users to increase their online engagement with particular content.

In a related study, Macskassy and Michelson [126] note that there was, and perhaps remains, a gap in research and extant literature relating to a detailed understanding of exactly what content is being spread through social media, and why. They also highlight that a lot of research at the time focused at the macro level of understanding online networks, and that there is also a need to look at the micro level. With the continuing development of social media, including the development of new platforms, there is likely to be a continued need to understand and re-understand exactly how such platforms are appropriated and used, both by organisations and by individuals. As such, it is likely that this identified ‘gap’ in research and understanding [126] will continue for some time, as new platforms and ways of using these platforms are introduced.

Organisations will often make use of social media to encourage the dissemination of some form of marketing message, such as particular products, new physical retail locations, upcoming events, or important announcements. In order to disseminate these to as wide an audience as possible, it is important to understand how and why individuals may share information with their own networks, circles of friends, or online followers. Many platforms encourage and facilitate this through features such as the ‘retweet’ function on Twitter, or the ‘share’ function on Facebook.

In research focusing on what may impact an individual’s decisions around what to share, Sharma and Cosley [171] highlight that, in a broad sense, factors such as demographics and cultural norms may play an important role. They develop this point further, noting that on an individual level, selecting what information to share further ultimately comes down to a combination of their preference for that content (however that preference may be formed), and what is, or what they deem to be, salient at that moment in them. One point that is raised by Sharma and Cosley [171] that still needs to be addressed is the predictability of individuals’ online sharing behaviours - if indeed they do share content at all [142].

This suggests, that while networks of followers around particular social media accounts may continue to grow numerically, there will remain a group of those followers (however large or small) that will very rarely, if ever, be encouraged to engage directly, or disseminate information from the account to their own network of followers. This highlights the need for organisations making use of social media to gain a detailed understanding of the make-up of their online audience, and to develop effective strategies for engaging with this audience. As each audience is likely to be made of up different individuals, it is unlikely that there will be one single solution or strategy that will be effective in all situations. Further to this, as audiences develop and grow (or decline), the individuals that make up that audience will also change - some will leave, others will join - and as such, whatever strategy or approach is adopted will have to be reviewed periodically to ensure that it still remains the most effective approach for that organisation at that moment in time.

While various strategies can be developed and implemented, it is important to note (particularly for those managing the social media accounts, and their superiors) that “*sharing is a voluntary process shaped by social forces such as people’s willingness to diffuse, attention to targets’ needs, and relations between sharer and target including tie strength and homophily*” [171].

This suggests that there is a need to understand users’ motivations for using social media, and through observations of their behaviour and network positioning, understand how they may be motivated to share and engage through social media.

3.5 Summarising Literature Review and Identifying Gaps in Extant Research

In Chapter 2 and Chapter 3, relevant research around social network analysis, and organisational use of technology (with a particular focus on marketing) has been discussed. In this section, the main focuses of this research is summarised, along with provocations from and 'gaps' in this research identified, as a means of motivating the work presented in this thesis, and potential future work. As the research presented here spans a significant period of time (especially given the fast-evolving nature of digital technologies), this section focuses on the field of research as it was when this research was undertaken. Relevant research conducted since this point is summarised and discussed in later chapters (see Chapter 9), with particular regard to the contributions of this research, limitations of the approaches undertaken here, and potential future work.

Drawing on the research discussed in Chapters 2 and 3, the following four main areas of provocation and 'gaps' in research are now explored in turn: understanding audience structure; attracting and retaining an audience; group identity, belonging, and behaviours; and content preferences, engagement, and sharing. Following this, the objectives of this research (as initially outlined in Section 1.2) are related to the identified gaps in research that motivated them. In later chapters of this thesis, such as Chapter 9, this is revisited, demonstrating how these research motivations and research questions are addressed by the studies presented here, and the resulting contributions.

3.5.1 Understanding Audience Structure

The need for organisations to understand, or deepen their understanding of, their audience or client base is not new and is not a unique research challenge in this area. However, consideration of aspects of prior work, and questions raised within that work - particularly regarding the underlying structure of the audience - is taken into account in the studies within this thesis.

At various points in the literature the need for understanding various aspects of the 'structure' of the audience, or network, has been raised. Some of these points were presented as open and unanswered questions, while at other times they have been discussed as important points that should be considered when conducting related research. For example, Wasserman and Faust [200] question whether social network metrics such as centrality are always important, or whether the importance of such metrics are context-dependent. In the context of online social networks and account followers, it is likely that some metrics will be more relevant than others, given the goals of the account owners. For example, centrality is likely to be more relevant when focusing on retaining followers.

Related to this, concepts such as bridges and cutpoints in networks may be useful identifiers to consider when identifying important or key nodes within a network [200]. In the context of online social networks, bridges and cutpoints could be used to identify users who are key in the process of disseminating information from one account to a much wider audience.

Highlighting the need for longitudinal studies and, in particular, how 'underlying mechanisms' play a role in the evolution of networks over time, Easley and Kleinberg [60] prove a key motivation for some aspects of the research presented here. This motivation is further supported by the work of Friedkin and Johnsen [73], who reiterate that there is a need to understand and 'grapple' with the actual structures of the underlying networks in such groups.

Understanding the structure of the audience, and in particular the underlying network, can also be related to the concept of market segmentation [155] which was discussed in detail in Section 3.1. An understanding of the structure of the network, and the communities within that network, could be used as a means of segmenting the audience, or supplementing more traditional means of grouping and segmenting an audience.

One consideration raised by Rentschler [166], which underpins each of the four areas discussed here, and the overarching theme of the research in this thesis, is the need for organisations to move to, or further develop, a 'post-modern marketing focus', which not only includes market segmentation, but also means of developing an understanding of the needs, wants, attitudes and behaviours of their clients or audience, and how these may change and develop over time. The research presented here demonstrates means by which a greater understanding of social media audiences, and their behaviours, can be understood, and as such could be incorporated in some form into the day-to-day operations of organisations.

3.5.2 Attracting and Retaining an Audience

The previously discussed literature highlights that in order to successfully grow the client base or audience of an organisation, it is often of importance to not only attract new individuals, but to also retain those that have already become 'clients' - in whatever form that may take - such as customers, or social media followers [109, 110].

This relates directly to the concept of 'churn' modelling or prediction, where those individuals at risk of 'churning', or stopping being a customer, can be identified. While there are numerous examples of churn modelling in various contexts, such as telephone service providers, and other 'traditional' commercial settings, at the time this research was first proposed and designed, there was not a great deal of literature surrounding churn modelling in online social networks, particularly when considered in the context of retail locations, as is the focus here.

This concept of churn modelling can also be linked back to the previously discussed focus of the underlying structure of networks. First, the positioning of individuals within a network, and the various metrics associated with this, and their impact on the likelihood of individuals churning is an area that is yet to be fully explored in this context. Second, how the changes in such positioning and metrics over time can also be used in the modelling of potential churn behaviours is another area that requires further investigation in this context.

The actions that might be taken by organisations, once those individuals that are at risk of churning have been identified, will differ from organisation to organisation, and is not the main focus of the work explored here. However, one further consideration that may be taken is how such information may be legitimately combined with other data, such as market segments (as discussed previously in Section 3.1).

3.5.3 Group Identity, Belonging, and Behaviours

While the focus of earlier social network theories and associated studies related to the notion of identities and belonging to a group or groups, this does not necessarily easily translate to less tangible online 'groups' of individuals - such as groups of followers of a particular social media account. While offline groups and identities might be more easily associated with due to a more tangible presence, how much individuals associate groups, group behaviours, and associated concepts, with following a social media account is an area that is yet to be fully explored.

Friedkin and Johnsen [73] discuss self-categorisation theory, questioning how individuals see and categorise themselves, and the result that this self-categorisation may have on their behaviours. While this has been explored in many contexts, how users categorise themselves when following and interacting with the online presence of retail locations is another area of research that requires further investigation.

This categorisation of self also relates to the concept of a prototypical group member - how individuals may perceive the 'average' or prototype group member to behave [73]. Understanding this more fully, in this particular context, is still an area that requires further investigation, and would include multiple facets, such as: understanding if individuals perceive themselves to be part of a group or community by following and interacting with a particular account, or accounts; whether this influences their behaviour, if they perceive themselves as belonging to a group, do they have a notion of what the 'prototype' group member 'looks' and behaves like, and how does any such concept affect their own behaviour in this context.

3.5.4 Content Preferences, Engagement, and Sharing

The final main area of motivation and 'gaps' in research identified through the literature review relate to the engagement with, and sharing of, social media content, and the role that individual preferences for particular content may have in this process.

The main motivation within this grouping is provided by Sharma and Cosley [171], who state that there still remains a need to better understand the predictability of individuals' sharing behaviours. This also links back to the previously identified motivation from Easley and Kleinberg [60], who highlighted the need for more longitudinal studies in fields such as this.

Models of diffusion, which are discussed in Section 2.4, are highly applicable in social media research in this context. What still remains as an area requiring further research is the motivation, or motivations, that lead to individuals not only reading content online, but sharing and diffusing that information further [60]. Easley and Kleinberg [60] refer to this as 'activation' - the point at which individuals decide to carry out this sharing of information. This is strongly related to the work of Macskassy & Michelson [126] and Sharma & Cosley [171], who note a need to understand what is shared online and why, and that the act of sharing content is down to the preferences of the individual, and what is seen as salient at the time. As these preferences will differ from individual to individual, and context to context, what remains as an area requiring further investigation is means by which these preferences might be understood at scale.

3.5.5 Gaps in Research as Motivation for Research Objectives

The four previous sections (Sections 3.5.1 - 3.5.4) outline various motivations and identified 'gaps' in existing research that have motivated the structure and focus of the studies in this thesis.

Table 3.1 outlines how these motivations and gaps in research are addressed in the research objectives of this thesis, and those that remain unaddressed but form the discussion of potential future work, discussed later in this thesis.

Each of these research objectives is addressed by the studies presented in this thesis. A further developed version of this table is included in Chapter 9, where the various studies and contributions related to these objectives are discussed in more detail.

3.6 Summary

In this chapter, existing work and related literature has been outlined and discussed, providing a context for the research presented in this thesis. Marketing research has been discussed, including the organisational need to understand the target market and audience, with approaches such as market segmentation and the concept of market churn. With this need identified, organisational adoption of technology and digital systems was discussed, including social media.

The final sections of this chapter focused on drawing the findings of this literature review together, including aspects of social network analysis discussed in the previous chapter. In these sections, motivations from, and identified 'gaps' in, this research are noted, and their relationship to the research objectives of this thesis are recorded.

The next chapter discusses the methodological approaches taken within the studies that follow, including a discussion of relevant ethical and legal considerations, along with the ethical approval processes undertaken during this research.

Table 3.1: Table demonstrating the extent to which the research objectives within this thesis address the motivations and 'gaps' in research highlighted during the literature review.

Motivations and Identified 'Gaps' in Research	**Research Objectives**
The need for longitudinal studies, as identified by Easley & Kleinberg [60]. Facilitating the further development of organisations moving to a post-modern marketing focus [166]. Understanding what is being shared online [126].	RO1 - To develop an understanding of how retail locations make use of social media, including the range of content being shared and how this may affect the levels of engagement from their social media audiences.
The need for longitudinal studies, as identified by Easley & Kleinberg [60]. A need to understand the reality of underlying networks in such studies, as per Friedkin & Johnsen [73]. Facilitating the further development of organisations moving to a post-modern marketing focus [166]. Further understanding 'churn' in an online social media context. How individuals' position in a network may effect churn.	RO2 - To develop a greater understanding of how online social media audiences may grow (or decline) and develop over time.
The relative importance of social network metrics (when used in context), as raised by Wasserman & Faust [200]. A need to understand the reality of underlying networks in such studies, as per Friedkin & Johnsen [73]. Facilitating the further development of organisations moving to a post-modern marketing focus [166].	RO3 - To investigate the extent to which social network analysis techniques and metrics can be used to indicate the likely organic growth or decline of social media audiences.
The need for longitudinal studies, as identified by Easley & Kleinberg [60]. Facilitating the further development of organisations moving to a post-modern marketing focus [166]. Further understanding 'churn' in an online social media context. Understanding what is being shared online [126]. Understand the predictability of sharing behaviours, and preferences for sharing [171].	RO4 - To implement and evaluate a method of generating profiles of social media users' engagement with specific content, to enable retail locations to understand the aggregate behaviour of their online audiences
The need for longitudinal studies, as identified by Easley & Kleinberg [60]. Facilitating the further development of organisations moving to a post-modern marketing focus [166].	RO5 - To develop an understanding of the role of engagement with social media content in the growth and development of social media audiences.
How the underlying structure of networks can be used to successfully augment other means of market segmentation. Do users perceive themselves as belonging to a 'group' when following a social media account, and how this effects their behaviour. Self-categorisation theory online - how users perceive themselves & how this may influence their behaviour. How users perceive the 'prototype' social media follower, and how this may influence their behaviour online.	*These are not the focus of the work presented in this thesis. These are discussed again later in the context of potential future work.*

Chapter 4

Methodology

4.1 Introduction

This chapter describes the methodological choices and approaches taken within this thesis. First, in Section 4.2, the aims and objectives of the research are revisited. Following this, in Sections 4.3 and 4.4, various approaches and considerations to qualitative and quantitative analysis are described. Following this discussion, in Section 4.5, the most appropriate approaches to this research are described and justified. Section 4.6 then describes the various ethical guidelines, legal frameworks, and some of the relevant academic literature which was consider throughout the design and implementation of this research; this section also includes a description of the ethical approval processes undertaken at both the University of Lincoln, and Northumbria University. Following this, in Section 4.7, the detail of the methods employed in each of the four studies are detailed.

4.2 Overview of Aim and Objectives

In Chapter 1, the main aim of the research presented in this thesis was "*to develop a greater understanding of social media users' engagement behaviours, and the effect that this engagement may have on maintaining and growing online social media audiences*". To meet this aim, five objectives have been identified. These objectives are designed to facilitate the overall aim of the thesis, are motivated by ongoing work with an industry partner, and provide a structured approach to the design of the relevant studies. The five identified objectives are listed below.

1. To develop an understanding of how retail locations make use of social media, including the range of content being shared and how this may affect the levels of engagement from their social media audiences.

2. To develop a greater understanding of how online social media audiences may grow and develop over time.

3. To investigate the extent to which social network analysis techniques and metrics can be used to indicate the likely organic growth or decline of social media audiences.

4. To implement and evaluate a method of generating profiles of social media users' engagement with specific content, to enable retail locations to understand the aggregated behaviour of their online audiences.

5. To develop an understanding of the role of engagement with social media content in the retention, growth and development of social media audiences.

Through the studies presented in this thesis, each of the objectives are addressed. In the next sections of this chapter, various approaches to data analysis are outlined and considered, before appropriate approaches for the purposes of this research are discussed.

4.3 Qualitative Research

In this section, qualitative research is discussed. First, in Section 4.3.1, qualitative research is defined, with its key concepts discussed, with particular attention paid to how such an approach could be applied and used within the context of this thesis. Following this, in Section 4.3.2, specific techniques and approaches to qualitative analysis are discussed in turn.

4.3.1 Introduction

Qualitative research is defined by Creswell and Creswell [49]:

> *Qualitative research is an approach for exploring and understanding the meaning individuals or groups ascribe to a social or human problem. The process of research involves emerging questions and procedures, data typically collected in the participants' setting, data analysis inductively building from particulars to general themes, and the researcher making interpretations of the meaning of the data.*

This definition highlights the applicability of such approaches to social media data, particularly relating to how individuals act and behave on such platforms, as the data is collected *"in the participant's setting"*, i.e. from the behaviour online, rather than in controlled experimental conditions. Yin [213], however, steers away from providing a strict definition of the exact nature and bounds of qualitative research:

> *Unfortunately, the breadth of what is called qualitative research, because of its relevance to different academic disciplines and professions, challenges you (or anyone) to arrive at a succinct definition.*

Instead, five underlying principles to qualitative research are provided by Yin:

> *1. Studying the meaning of people's lives, in their real-world roles.*
> *2. Representing the views and perspectives of people in a study.*
> *3. Explicitly attending to and accounting for real-world contextual conditions.*

4. Contributing insights from existing or new concepts that may help to explain social behaviour and thinking.
5. Acknowledging the potential relevance of multiple sources of evidence rather than relying on a single source alone. [213, p.9]

These five points, when considered individually and collectively, again demonstrate that qualitative approaches are suitable in addressing, to varying extents, the objectives of the research, previously outlined in Section 1.2 and Section 4.2. Qualitative research is useful for the studying of people's lives, as experienced under real-world conditions, with people behaving in their regular, everyday roles and habits, without impact or influence from any experimental setting [213, p.10]. Further, qualitative research explicitly embraces the context and conditions under which the actions of individuals take place [213, p.10]. As such, this demonstrates that qualitative research is a suitable means of understanding behaviour on social media. Within the context of this thesis, it can be used in addressing objectives 1, 4, and 5.

There are many approaches through which qualitative research can be approached and undertaken, a selection of the most relevant approach is discussed in the following section.

4.3.2 Approaches to Qualitative Data Analysis

There are multiple approaches to qualitative analysis, each suited to different types of research questions, and particular types of data. Braun and Clarke [28] focus on four of these approaches: thematic analysis, interpretative phenomenological analysis, grounded theory lite, and pattern-based discourse analysis. These four approaches are summarised in Table 4.1, and are discussed in turn in the following sub-sections. Later, in Section 4.5, the chosen approaches to data analysis within this thesis are detailed and justified.

Table 4.1: Various qualitative methods, and the research questions and data types to which they are best suited. Adapted from Braun and Clarke (2013) [28, p.50].

Method	**Suitable Types of Research Question**	**Suitable Types of Data**
Thematic Analysis	Any, except language practice	Any, no ideal data type
Interpretative Phenomenological Analysis (IPA)	Experience, understandings, perceptions	Interviews are ideal. Researcher-directed diaries, qualitative surveys and focus groups are also used.
Grounded Theory Lite	Any, except language practice. Influencing factors is ideal.	Any, although interviews are common.
Pattern-based Discourse Analysis	Accounts of practice, representation, construction.	Any, no ideal data type.

4.3.2.1 Thematic Analysis

Thematic 'coding' of content (see Glossary of Terms and Definitions), in some form, is common across many methods and approaches to qualitative data analysis. Thematic analysis focuses on understanding the content of data through the application of 'codes' that label this content, and the eventual grouping of these codes into broader 'themes'. Thematic analysis techniques are often divided into three broad clusters, which have been described as 'coding reliability', 'codebook', and 'reflexive'. Each of these clusters of techniques are best employed in different contexts, with 'coding reliability' techniques focusing, for example, on the use of multiple coders to ensure high levels of agreement and reliability between these coders. 'Codebook' thematic analysis focuses on techniques that apply existing codebooks (i.e. from prior research), to datasets. With the earlier works of Braun and Clarke often cited [27], their approach to thematic analysis has been (broadly speaking) widely adopted across a range of research areas. However, in a more recent publication [27] Braun and Clarke seek to dispel some of the misunderstanding and misapplication of their approach to thematic analysis. Braun and Clarke now refer to their approach (and similar approaches proposed by others, such as [114]) as *reflexive* thematic analysis, to *"emphasise the active role of the researcher in the knowledge production process"* [29, p.6].

As two of these three groups of approaches (coding reliability, and codebook) generally require multiple coders and researchers, or the application of existing coding schemas, these were identified as not being the most suitable approach to thematic analysis to consider, in this context. As such, reflexive thematic analysis was identified as the most appropriate potential approach from these three. The importance of the researcher, or coder, will be discussed again in later chapters, with particular reference to the domain-specific knowledge of those already situated within the organisations or domains being studied. Reflexive thematic analysis, as described by Braun and Clarke [29], follows six stages: data familiarisation, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and reporting. Each of these stages are outlined below, before the relevance of such an approach to the analysis of social media is discussed.

Familiarisation with Data

The first stage of the process - familiarisation with data - allows for the researcher to gain an understanding of the range of content in the dataset. This is particularly important in cases where the researcher has not necessarily collected the data through interactive means, or is using a secondary dataset [27]. The aim of this initial stage is to become '*intimately familiar*' with the content of a dataset, and to begin to take note of elements of the data that may be relevant to the research question [28], looking for what is interesting, rather than attaching formal labels at this point [30]. These '*codable moments*' [23] that attract the researcher's attention can then be revisited more fully in the next phase of analysis.

Generating Initial Codes

Once familiar with the breadth of data contained within the dataset, it is then necessary to 'code' the dataset - that is, to identify "*aspects of the data that relate to your research questions*" [28, p.206]. At this point, the approach taken can often be categorised in one of two ways, which Braun and Clarke term '*selective*' or '*complete*' coding. Whilst the former involves identifying a set of instances of phenomena in which you're interested, the latter involves identifying any aspect which may prove to be of interest in relation to the research question(s). Regardless of the chosen approach, the aim of the coding process remains the same:

> *The aim of coding and theme development in reflexive thematic analysis is not to "accurately" summarise the data, nor to minimise the influence of researcher subjectivity on the analytic process, because neither is seen as possible nor indeed desirable. The aim is to provide a coherent and compelling interpretation of the data, grounded in the data. The researcher is a storyteller, actively engaged in interpreting data through the lens of their own cultural membership and social positioning, their theoretical assumptions and ideological commitments, as well as their scholarly knowledge. [30, p.6]*

Searching for Themes

Identifying and developing themes from the data is an active process, rather than the passive process suggested when themes are mis-described as 'emerging' from the data [28]. Whereas the codes generated in the previous phase of analysis capture a single point or idea, themes should be conceptual, and organise groups of codes around a central concept.

In this phase of analysis, codes are grouped and collated, with the aim being to identify similarity and overlap between them. These themes should identify and "*capture the most salient patterns in the data relevant to answering your research questions*" [28, p.231].

Reviewing Themes

This phase can often be thought of as a quality control phase, "*checking to determine whether your candidate themes fit well with the coded date, and the dataset you collected*" [28, p.233]. This phase involves reviewing the current themes in relation to the data that falls within each theme, ensuring that the theme accurately describes the data. In cases where the theme does not accurately summarise the data, it may be necessary to rename themes, create multiple themes in place of a single original theme, or combine themes together. By the end of this stage, themes should be coherent (have a central organising concept), distinctive (from other themes), work together, and relate to the research question(s) in some way.

Defining and Naming Themes

Building on the previous phase, this phase focuses on developing the analysis, scope, and focus of each theme - determining 'the story' of each theme [27]. It is during this stage that the names used to describe each theme are decided.

Reporting

This final phase, which may take different forms depending on the audience and venue for reporting, "*involves weaving together the analytic narrative and data extracts, and contextualising the analysis in relation to existing literature*" [27].

Potential Relevance to this Research

Thematic analysis, and in particular *reflexive* thematic analysis is a suitable approach for this research, as it allows for a structured understanding of the content of text-based content (such as social media content). Further, while some approaches to thematic analysis require the involvement of multiple coders or researchers, or the application of existing codebooks, reflexive thematic analysis focuses and emphasises the active role of the researcher in the knowledge production process. In addition to this, the approaches within reflexive thematic analysis can be applied to multiple content types, either in combination or isolation, and as such, again, makes it suitable for use with social media data, that could take the form of text, still images, video, audio, or combinations thereof.

4.3.2.2 Interpretative Phenomenological Analysis (IPA)

Developed by Smith and colleagues, IPA primarily focuses on how individuals make sense of their lived experiences. IPA (see Glossary of Terms and Definitions) can be employed to analyse the experiences of a single participant, or to generate themes across a group of participants [28]. The phenomenological aspect of IPA concerns how individuals make sense of lived experiences, with the interpretative aspect relating to how this 'sense making' is achieved through the researcher's interpretative activities.

When undertaking IPA, "*researchers need first of all to access rich and detailed personal accounts. These accounts will be elicited from persons who are able and willing to offer us a view of the phenomena under investigation*" [178, p.40]. Compared to the previously discussed reflective thematic analysis, IPA involves a more active participation from study participants, with researchers involving participants in the 'sense making' of the topic or topics being researched. As Smith, Flowers, and Larkin write: "*In IPA we are assuming that our data can tell us something about people's involvement in and orientation towards the world, and/or about how they make sense of this*" [178].

Potential Relevance to this Research

While IPA may be relevant to other areas of research surrounding social media, it is not necessarily best suited to the overall approach of the research presented within this thesis. Due to the nature of IPA, which requires working closely with research participants, such an approach is more suited to studies which involve (for example) understanding individual users' experiences and motivations for social media use, rather than understanding aspects of behaviour of large numbers of people.

4.3.2.3 Grounded Theory Lite

Grounded Theory, and in particular Grounded Theory Lite, has some aspects in common with the previously discussed Reflexive Thematic Analysis. As defined by [28, p.80], Grounded Theory Lite:

> *Aims to generate a taxonomy of categories (clusters of related codes; similar to themes in TA and IPA) from data, with some indication of the relationships between concepts and the relative importance of concepts to the research question.*

Providing a more concrete definition of Grounded Theory, Strauss and Corbin describe it as "*a qualitative research method that uses a systematic set of procedures to develop and inductively derived grounded theory about a phenomenon*" [184, p.24]. Whereas approaches such as reflexive thematic analysis provide a means to analyse data, Grounded Theory goes further, providing a framework through which substantive theories can be derived and developed.

Potential Relevance to this Research

Grounded Theory Lite is one approach that could be considered appropriate and relevant to the aims and objectives of the research in this thesis. It provides a framework for analysis of data, through to the development of a substantive theory. It is also noted that this approach is suitable for use with any data type, although interviews are often seen as the most common data source [28, p.50].

4.3.2.4 Pattern-Based Discourse Analysis

Described by Braun and Clarke (2013), pattern-based discourse analysis methods are "*concerned with patterns in language use connected to the social produce of reality, and with understanding how accounts of objects and events are constructed in particular ways*" [28, p.177]. While many different varieties of pattern-based discourse analysis exist, including thematic discourse analysis, post-structuralist discourse analysis, interpretative repertoires, and critical discursive psychology [28], each of these tends to focus on the discursive feature of language, including power relationships and the functions of discourse. Broadly speaking, approaches to discourse analysis focus on what language does, rather than 'simply' on what language is used [28, p.187].

Potential Relevance to this Research

The main focus of discourse analysis goes beyond the aims of the qualitative analysis aspects of the research presented in this thesis. While further studies in this, and related, areas may make use of discourse analysis, such an approach is not necessarily the most suitable to meet the aims and objectives of this research.

4.4 Quantitative Data Analysis and Descriptive & Inferential Statistics

In the previous section, Section 4.3, various approaches to qualitative data analysis were outlined, and their relevance to the analysis of social media content discussed, within the context of this thesis. In this section, various key elements of quantitative analysis and inferential statistics are considered, again within the context of the research presented in this thesis.

4.4.1 Introduction

Describing quantitative analysis, Walliman writes that "*quantitative analysis deals with data in the form of numbers and uses mathematical operations to investigate their properties*" [199, p.131]. Lindgren develops this further, describing how quantitative research is "*characterised by the deductively oriented statistical study of large numbers of cases*" [118, p.29]. With quantitative analysis of large datasets a suitable means of addressing the aims and objectives previously discussed in Sections 1.2 and 4.2, the following sections outline key aspects of quantitative analysis and inferential statistics.

4.4.2 Hypothesis Testing

The testing of hypotheses, statements that have yet to be tested against collected data [199], in the context of this thesis, allows for testing relationships between metrics (outlined in Section 1.2). Two forms of hypotheses, null and alternative, are used. While the alternative hypothesis describes a hypothesised relationship between the variables being tested, the null hypothesis describes the absence of effects between those variables.

The purpose of the two forms of hypothesis is to find evidence (at an appropriate level of significance) to reject or falsify the null hypothesis (often indicated as H_0). If the null hypothesis - and therefore the lack of effect between the variables - can be confidently rejected, then the alternative hypothesis, describing the effect between the variables - can be accepted.

4.4.3 Type I and Type II Errors

Two types of errors need to be considered when conducting analyses such as the statistical analyses used within this thesis, Type I errors and Type II errors. Type I errors, often referred to as 'false positives' occur when a null hypothesis is rejected that is actually true within the population being studied. Type II errors, referred to as 'false negatives', occur when a false null hypothesis is not rejected [13].

4.4.4 Significance Levels

Often used synonymously with the term 'p value', the statistical significance value indicates the probability that the reported result may be due entirely to chance. Though there is no explicit level at which statistical significance is deemed to be sufficient in all circumstances, it is common practice to treat p values of less than 0.05 as sufficiently significant. Such a result would indicate that the results being reported are erroneously accepted in fewer than 5% of cases.

In the context of hypothesis testing, as used within this thesis, the null hypothesis (indicating no relationship between the variables being studied) would be rejected where results have an associated p value below this pre-determined level, referred to as the alpha (α) value. Combining this with the definition of Type I errors, described previously, an alpha value of 0.05 would indicate that it has been deemed acceptable to have a 5% chance of incorrectly rejecting a true null hypothesis. Care must also be taken when interpreting results when multiple tests are conducted, as the likelihood of Type I errors occurring is increased in such a situation, with the likelihood increasing as the number of tests conducted increases.

4.4.5 Parametric and Non-Parametric Tests

Statistical procedures can be grouped into two categories: parametric and non-parametric (see Glossary of Terms and Definitions). While parametric procedures rely on assumptions about the distribution in the underlying population and data, non-parametric procedures rely on no such assumptions. A summary of common types of analysis, and the parametric and non-parametric equivalent procedures for these types of analyses can be found in Table 4.2.

Table 4.2: Examples of parametric and non-parametric procedures for common types of analysis. Adapted from [95].

Type of Analysis	Parametric Procedure	Non-parametric Procedure
Compare means between two distinct or independent groups	Two-sample t-test	Wilcoxon rank-sum test
Compare two quantitative measurements taken from the same individual	Paired t-test	Wilcoxon signed-rank test
Compare means between three or more distinct or independent groups	Analysis of variance (ANOVA).	Kruskal-Wallis test
Estimate the degree of association between two quantitative variables	Pearson coefficient of correlation	Spearman's rank correlation

While Table 4.2 summarises different procedures for common types of analysis, it is important to note that the use of parametric tests with data that does not meet the assumptions of parametric tests could lead to erroneous results. In these instances, where assumptions cannot be made or met, it is more appropriate to use non-parametric tests. Decisions regarding which tests are appropriate in the context of the studies presented within this thesis are discussed in the relevant sections of this chapter - Sections 4.7.1, 4.7.2, 4.7.3, and 4.7.4.

4.5 Appropriate Approaches for this Research

In the previous sections, various approaches to qualitative data analysis have been outlined, along with various aspects of quantitative analysis and inferential statistics. In this section, the approaches selected for use within the research presented in this thesis are detailed and justifications for this selection are provided. In the following sub-sections, a mixed methods approach is first outlined (in Section 4.5.1); following this, in Section 4.5.2, the selected approach to qualitative data analysis is discussed; finally, in Section 4.5.3, various aspects of quantitative analysis and inferential statistics are detailed.

4.5.1 A Mixed Methods Approach

Based on the aims and objectives of the research presented here, it is necessary to implement both qualitative and quantitative approaches to data analysis. Qualitative approaches will facilitate an understanding of the content of social media posts (required for objectives 1, 4, and 5), while quantitative approaches will contribute toward some aspects of each of the five stated objectives. This mixed methods approach facilitates the completion of the previously outlined objectives of this research, and allows for the use of multiple approaches, each with their own focus and relative strengths.

> *Today, however there is an increasingly widespread consensus that the employment of combinations of 'qualitative' and 'quantitative' methods is a valid and recommended*

strategy, which allows researchers to benefit from their various strengths, and balance their respective weaknesses. [118, p.29]

While Lindgren [118] outlines the potential benefits of mixed methods research on a general level, there are many examples of studies relating to social media data that have employed a mixed methods approach to data analysis [181]. Of the studies analysed by Snelson in a 2016 literature review [181], those which included 'big data' (e.g. large datasets collected from social network sites) almost exclusively featured a mixed methods research design. From the 55 mixed methods studies analysed, the underlying analytic approaches included coding, content analysis, and statistical analysis, which is similar to the approaches employed in this thesis.

4.5.2 Qualitative Analysis of Social Media Content

In Section 4.3 various approaches to qualitative data analysis were outlined, along with their potential relevance to the studies presented in this thesis. Based on the focus and relative merits of each of these approaches, it was determined that reflexive thematic analysis was an appropriate choice for the qualitative data analysis aspects of this thesis.

The qualitative aspects of the analyses required in this thesis relate specifically to understanding the content of social media posts, which reflexive thematic analysis facilitates. Further, reflexive thematic analysis can be used appropriately by a single researcher, which differs from other variants of thematic analysis. While other discussed approaches, such as discourse analysis and interpretative phenomenological analysis, would be appropriate for future related research in this field, the frameworks they provide go beyond what is required in this context. The use of reflexive thematic analysis also has further benefits, such as being flexible in terms of what types of content it can be applied to. While the research presented here focuses on text-based content, the same process can be replicated on visual, audio, and mixed-media content.

As such, qualitative analysis of social media posts in the research within this thesis will follow the six-stage reflexive thematic analysis approach described by Braun and Clarke [28, 29, 30]. The six stages are described earlier, in Section 4.3.2.1, and their specific application within each study is described in the relevant sections - Sections 4.7.1, 4.7.2, 4.7.3, and 4.7.4.

4.5.3 Hypothesis Testing and Quantitative Analysis

In order to meet the objectives of this research, quantitative methods and hypothesis testing will be employed within the studies. The specific details of each study are provided in Sections 4.7.1, 4.7.2, 4.7.3, and 4.7.4. Where relevant, these sections will include details regarding the justification and use of parametric or non-parametric tests, the testing of hypotheses, and the use of appropriate significance levels.

4.6 Ethical, Legal, and Best Practice Considerations

Various ethical and legal guidelines and frameworks are relevant and applicable when considering the collection and analysis of social media data. In this section, relevant ethical frameworks and guidelines are discussed (in Section 4.6.1), before legal requirements and considerations are outlined (in Section 4.6.2). Following this, what is considered as 'best practice' in academic literature is discussed (in Section 4.6.3), followed by an outline of various social media platforms' terms and conditions, and how these have developed and changed over time (see Section 4.6.4). Section 4.6.5 then applies these guidelines, frameworks, laws, and terms and conditions to the research within this thesis. Finally, Section 4.6.6 details the ethical approval processes undertaken during this research.

4.6.1 Ethical Frameworks and Guidelines

Various ethical frameworks and guidelines now exist and have been introduced and developed by organisations during the course of this research. Here, those frameworks and guidelines are outlined. While this set of guidelines is not intended to be exhaustive, these are the guidelines (and their future revisions) that were consulted and adhered to during the course of this research.

4.6.1.1 Association of Internet Researchers (AoIR)

The Association of Internet Researchers (AoIR) was founded in 1999, and has (to date) provided 3 major guideline documents related to internet-based research, published in 2002, 2012, and 2019.

Version 1.0 , Published 2002
Their first publication [7] was released in 2002, before platforms such as Facebook, Twitter, and Instagram were developed. As such, these guidelines do not necessarily reflect the volumes of data that can now be acquired through the APIs of these platforms. However, these guidelines do highlight the general issues that are faced when considering ethical approaches to online and social media research:

> *online researchers may encounter conflicts between the requirements of research and its possible benefits, on the one hand, and human subjects' rights to, and expectations of, autonomy, privacy, and informed consent.*

> *as online research takes place in a range of new venues ... researchers, research subjects, and those charged with research oversight will often encounter ethical questions and dilemmas that are not directly addressed in extant statements and guidelines.*

> *Especially as Internet research may entail a literally global scope, efforts to respond to ethical concerns and resolve ethical conflicts must take into account diverse national and cultural frameworks.*

These statements highlight key aspects that should be considered, and have been considered as this research progressed, namely: balancing requirements and expectations, adapting to situations that are not directly addressed by existing guidelines, and taking into account the range of national and cultural frameworks and viewpoints.

Version 2.0 , Published 2012

Version 2.0 of the AoIR guidelines [8], was the most recently published AoIR guidelines document available when this research was initially planned and undertaken. As such, these guidelines formed a key consideration when designing the studies and applying for relevant ethical clearance from the university ethics boards.

Building on the previous report, this set of guidelines highlights that

> *... no set of guidelines or rules is static, the fields of internet research are dynamic and heterogeneous. This dynamism is reflected in the fact that as of the time of this writing, no official guidance or 'answers' regarding internet research ethics have been adopted at any national or international level.*

Demonstrating this development since the first report was released ten years earlier, this second version of the guidelines begins to move towards a more explicit focus on the notions of potential vulnerability and harm to those whose data is included within the research datasets. While the nature and definition of 'human subjects' *"no longer enjoys the relatively straightforward definitional status it once did"* [8], focusing on such a definition may not be as relevant as considerations around harm, vulnerability and personally identifiable information.

The studies in this thesis were designed with these three considerations in mind. By focusing on content shared from the social media accounts of organisations, and by developing a means of understanding social media engagement at an aggregate, rather than individual level. the risk of harm to individual was reduced. These considerations are discussed in greater detail in Section 4.6.5.4.

Version 3.0 , Published 2019

The most recent version of the AoIR guidelines was published in 2019 [9], after the research in this thesis was completed, and thus did not form part of ethical considerations for these studies. However, the continued development in the social media landscape is evident within this third version of the guidelines. These now include a greater focus on the notion of 'big data' and the implications this poses for gaining informed consent, and protecting the researcher(s), particularly in the wake of events such as 'gamergate' [187].

While these guidelines were not available while the research presented here was carried out, these guidelines are not contrary to the choices made when conducting these studies. They do, however, once again highlight that ethical considerations and guidelines need to change and develop as the field they relate to changes and develops.

4.6.1.2 British Psychological Society (BPS)

At present, there are two versions of 'Ethics Guidelines for Internet-mediated Research', published by The British Psychological Society; the first of these was released in 2013 [32], with the second version published in 2017 [33]. As such, both of these were relevant during the studies presented in this thesis. The first version was considered when the studies were first designed, and the second version acted as a further means by which to review the initial decisions made, during the course of the research.

Ethics Guidelines for Internet-Mediated Research, Published 2013

This initial set of guidelines lays out four main principles to consider "*when designing, implementing, or assessing an IMR [internet-mediated research] study*" [32]. These four principles include: respecting the autonomy and dignity of individuals, scientific value (of the research), social responsibility, and maximising benefits while minimising harm. None of these four principles are unique to this set of guidelines, but by summarising the underlying ethical principles into four main categories, it highlights the core issues that should be considered when designing and implementing any internet-mediated research study.

The first principle, 'Respect for the autonomy and dignity of persons' includes the topics of public/private spaces online, as well as the need (where applicable and feasible) to gain informed consent from study participants. Regarding these two topics, this set of guidelines provides useful guidance that was used during the design of these studies.

> *..observation of public behaviour needs to take place only in public situations where those observed 'would expected to be observed by strangers'.*

> *When there is a level of ambiguity concerning whether data are 'in the public domain' or not, researchers should particularly consider the extent to which undisclosed observation may have potentially damaging effects for participants, before making decisions on whether to use such data and whether gaining valid consent is necessary.*

While the application of these principles to this work is outlined more clearly in later sections of this chapter, these two points are adhered to by focusing on the content shared on the public accounts of organisations, and by developing a means of understanding engagement behaviour at an aggregate, rather than individual level.

The third principle listed in this set of guidelines, that of 'social responsibility' also covers (again) the debate on the distinction between public and private spaces online, as "*intrusions from researchers into spaces considered private by their users may be invasive, unwelcome and socially irresponsible*" [32, p.16].

The fourth principle, 'maximising benefits and minimising harm' is particularly relevant to some aspects of the design of the studies presented in this thesis, specifically the use and inclusion of quoted content from social media posts.

> *Serious consideration should be given to whether publishing such traceable quotes requires specific valid consent from the individual, and it should be avoided in any cases where possible consequential risk and harm to participants is non-trivial.*

While all quotes used within these studies are from public organisations, rather than individuals, such considerations were made when constructing the original research questions and designing the studies. This question around the need to include verbatim quotes from individuals in publications and other outlets can also be seen in academic literature, which is summarised in a later section of this chapter, Section 4.6.3. Many of these examples take the same approach that is suggested later in the BPS (2013) guidelines:

Some researchers have addressed this issue by suggesting paraphrasing or combining quotes used in publications, and this could be considered if it is consistent with the research design.

As outlined at the beginning of this section, the ethical standpoints taken within this set of guidelines are not unique, but the repetition of the core concerns serves to reinforce the core aspects that need to be considered and adhered to in such contexts. The application of these guidelines to the studies presented here is discussed more fully in Section 4.6.5.

Ethics Guidelines for Internet-Mediated Research, Published 2017

In 2017, an updated version of 'Ethics Guidelines for Internet-mediated Research' was released by the British Psychological Society [33]. The structure and content of these guidelines was, in some ways, quite similar to those published in 2013 [32], with some updates made that reflect the changes in the social media landscape in the time elapsed between the publication of the two documents.

While the four main ethical principles remained the same, some content has been updated to reflect developments in the social media and online landscape. Updates to the first principle now reflect the potential opportunities for profiling both individuals and groups that sheer scale of available data now affords. This is reflected in the updated name of the first principle, which is now 'Respect for the Autonomy, Privacy and Dignity of Individuals and Communities', and discussion of the "*unobtrusive collection of very large data sets, involving the traces of people's online behaviours*" [33].

The publication of these guidelines served as a further opportunity to reassess the research from an ethical standpoint during the course of the research studies. As the fundamental nature of the guidelines had not changed, this did not raise any issues that had not been addressed, mitigated, or considered from the outset.

4.6.1.3 British Sociological Association (BSA)

The British Sociological Association published ethical statements [34], guidelines [38], and case studies [35, 36, 37, 39, 40, 41] in 2017. While these documents were published after the design and implementation of the majority of the studies presented here, the release of these documents provided a further opportunity for reviewing and assessing the decisions that had been made, as well as any decisions that were yet to be made regarding publication of data etc.

The 'Ethics Guidelines and Collated Resources for Digital Research' [38] serves as an annex to the main BSA 'Statement of Ethical Practice' document [34]. This document, rather than outlining a specific set of guidelines that should be followed, focuses more on summarising other work and guidelines to provide an overview of the commonly accepted ethical practice in given contexts, as well as the legal requirements.

One particularly relevant aspect of these guidelines is the summarisation of contexts in which informed consent may not be necessary or required for data collection. Relevant to the context of this thesis is the notion of the online 'public' space; summarising other guidelines and frameworks, the document again reiterates that informed consent is often not deemed to be required if the data is collected where people would reasonably expect to be observed by strangers. With this in mind, the use of social media content created and shared by organisations is listed in one of the case-study documents as a 'low risk' activity [39], as it is often clear that organisations are creating this content and sharing it with the intention of it being public and reaching a wide audience.

Again, relating to the publication of quotes, another case study document [35] highlights that anonymising quotes, including paraphrasing them, is an approach which is often used, if and when it is felt necessary to include examples of social media content in reports and publications. The implications of this are discussed in more detail in Section 4.6.5.

4.6.1.4 Development and Changes to Frameworks and Guidelines

As can be seen in the previous sub-sections, these ethical frameworks and guidelines both highlight the changing contexts in which internet-based research is conducted, but also are subject to changes and development themselves. Research in this thesis was conducted between 2013 and 2019; in this time the field of social media research has developed considerably, as have the associated ethical guidelines. As such, the ethical considerations directly related to this research (detailed in Section 4.6.5, and the ethical approval processes (Section 4.6.6) need to be considered in relation to the guidelines and institutional approval processes as they were at that time, and not necessarily in relation to the more developed and detailed guidelines and approval processes that may be in place now.

4.6.2 Laws and Regulations

Legal obligations and considerations will differ between countries and regions. In this section, relevant UK and EU laws are discussed in relation to social media research. Due to the timescale over which this research was conducted, various laws are, or were, applicable at varying points throughout the research presented in this thesis. First, the 1998 UK Data Protection Act [1], and its implications, is discussed. Following this, the EU General Data Protection Regulation (GDPR) [63] and its UK-based implementation - the 2018 Data Protection Act [2] are discussed.

4.6.2.1 Data Protection Act (1998)

The Data Protection Act (1998) [1] formed the primary legal considerations when planning and developing the research presented in this thesis, and its application to the research was included as part of the applications for ethical approval made at both the University of Lincoln, and Northumbria University.

This Act included 8 core principles, regarding the storage and processing of personal data. As social media data can, and often does, contain personal data, this act was relevant and applicable to the research conducted and presented here. These include processing data for a specified and lawful purpose; not collecting an excessive amount of data beyond what is required; ensuring the data is accurate and where necessary up to date; ensuring the data is not kept for longer than is necessary; and taking appropriate steps to prevent loss, destruction and improper access to the data. The steps taken to address these principles are detailed in Section 4.6.5. This Act was superseded by the Data Protection Act (2018) - the UK enactment of the EU General Data Protection Regulation, which is explored in more detail in the next section.

4.6.2.2 Data Protection Act (2018) and EU General Data Protection Regulation

The Data Protection Act (2018) [2] is the UK national law which complements the EU General Data Protection Regulation, and superseded the Data Protection Act (1998). As these regulations were not applied retrospectively, they did not apply to the data collection conducted before May 2018, and as such were not directly relevant to the data collection within the studies presented in this thesis.

However, if these studies were to be repeated, or built upon in the future, then the Data Protection Act (2018) would certainly be relevant and applicable, and would cover elements such as the processing of personal data, including data storage and retention.

4.6.3 Academic Literature and 'Best Practice'

As legal frameworks and ethical guidelines change and develop over time, so does the notion of 'best practice' as indicated in academic literature. Due to the relatively short period of time in which academic literature can be written and published, compared to professional guidelines and legal frameworks, academic literature can often respond to changes and developments much quicker. As discussed previously, the landscape of social media has changed dramatically over the past decade, and continues to develop. As such, laws, ethical guidelines, and academic best practice has developed too. It is important to frame the approaches undertaken in this research not just in terms of the 'state of the art' at present, but also in the context in which the research was originally designed and undertaken.

Table 4.3 summarises the main points from a selection of academic work, between 1998 and 2019. This list is not intended to be exhaustive, but instead provides an overview of relevant academic literature, and demonstrates how the main concepts and viewpoints have developed over time. The application of some of the points raised in these papers to the ethical approval processes undertaken for this research are then explored in later sections of this chapter.

The underlying narrative to academic literature over this time demonstrates both how the field of internet-mediated research has developed, but also how ethical concerns and major issues have become the focus of academic literature during this time. As social media platforms have been introduced, grown, and developed, so too has the depth and breadth of data made available to researchers through these platforms, and as such the focus of ethical considerations has changed and developed over time.

For example, earlier literature, such as in 1998 - 2000 raised concerns around the quantity of information that was being made available online, and how this might infringe the privacy of individuals. Following this, literature in the early 2000s focused on the application of extant ethical frameworks to online research, raising questions such as 'what online spaces are to be considered public spaces?'.

Following the launches of Facebook, YouTube, and Twitter between 2004 and 2006, the ethical focus began to move towards concerns around the notions of 'networked publics', as these platforms introduced not only data regarding individuals, but also focused on relationships between these individuals. This again raised questions around the concept of informed consent, as data no longer related to just individuals, but now also to the wider network of those around them.

As larger and larger datasets are made available through these platforms, examples of potential misuse of these datasets are being used to demonstrate how ethical codes of conduct and frameworks are still lacking with regard to social media research. Many papers after 2008 cite the 'Tastes, Ties, and Time' project as one such example, in terms of how data is collected, how consent is (or is not) required, how data is anonymised and made available to others, and how such studies should be reported.

More recent academic literature, published after 2013 (when this research commenced) continue to highlight the ongoing questions around informed consent in large-scale social media studies, as well as beginning to look at how social media users (as research participants, in this context) perceive the use of their data in research studies. Further to this, and potentially in light of changes in legislation, including the EU GDPR and associated laws, focus continues to move towards the issues around problematic profiling of individual social media users, for example, profiling based on 'protected characteristics', such as gender, sexual orientation, or religious beliefs. Figure 4.1 presents a timeline of relevant academic literature, legal frameworks, and ethical guidelines in relation to the time frame of the research presented in this thesis.

Section 4.6.5, influenced by legal requirements, ethical guidelines, and academic literature, outlines the key concerns and approaches applied to the research presented in this thesis.

4.6.4 Social Media Platforms' Terms & Conditions

The Terms & Conditions of social media platforms govern not only the relationship between platform and user, but also between the platform and those requesting access to the data they make available. As such, the relevant Terms and Conditions of the social media platforms were applicable, and adhered to during the data collection phases of these studies.

To adhere to the Terms and Conditions of both Facebook and Twitter, data was collected using the relevant APIs, with access restricted to data that was publicly available at that time. Since the point of data collection, however, social media platforms have continued to restrict and control access to this data, both in terms of the quantity of data that can be accessed within a given time frame, but also in terms of exactly what data can be accessed and by whom. Where relevant, specific access restrictions are included in the descriptions of each of the four studies, detailed in Section 4.7.

4.6.5 Application of Ethical and Legal Frameworks to this Research

Previous sections have included discussion around the various legal requirements, ethical guidelines, and academic literature relevant to the studies presented here. As part of the process of undertaking this research, various applications for ethical approval were undertaken, at both the University of Lincoln, and later at Northumbria University. The following sections highlight the approaches taken related to three main areas of ethical concern - informed consent (Section 4.6.5.1), respecting privacy by maintaining the anonymity of individuals (Section 4.6.5.2), and profiling the behaviour of social media users (Section 4.6.5.3). Following this, platform-specific ethical considerations are discussed in Section 4.6.5.4.

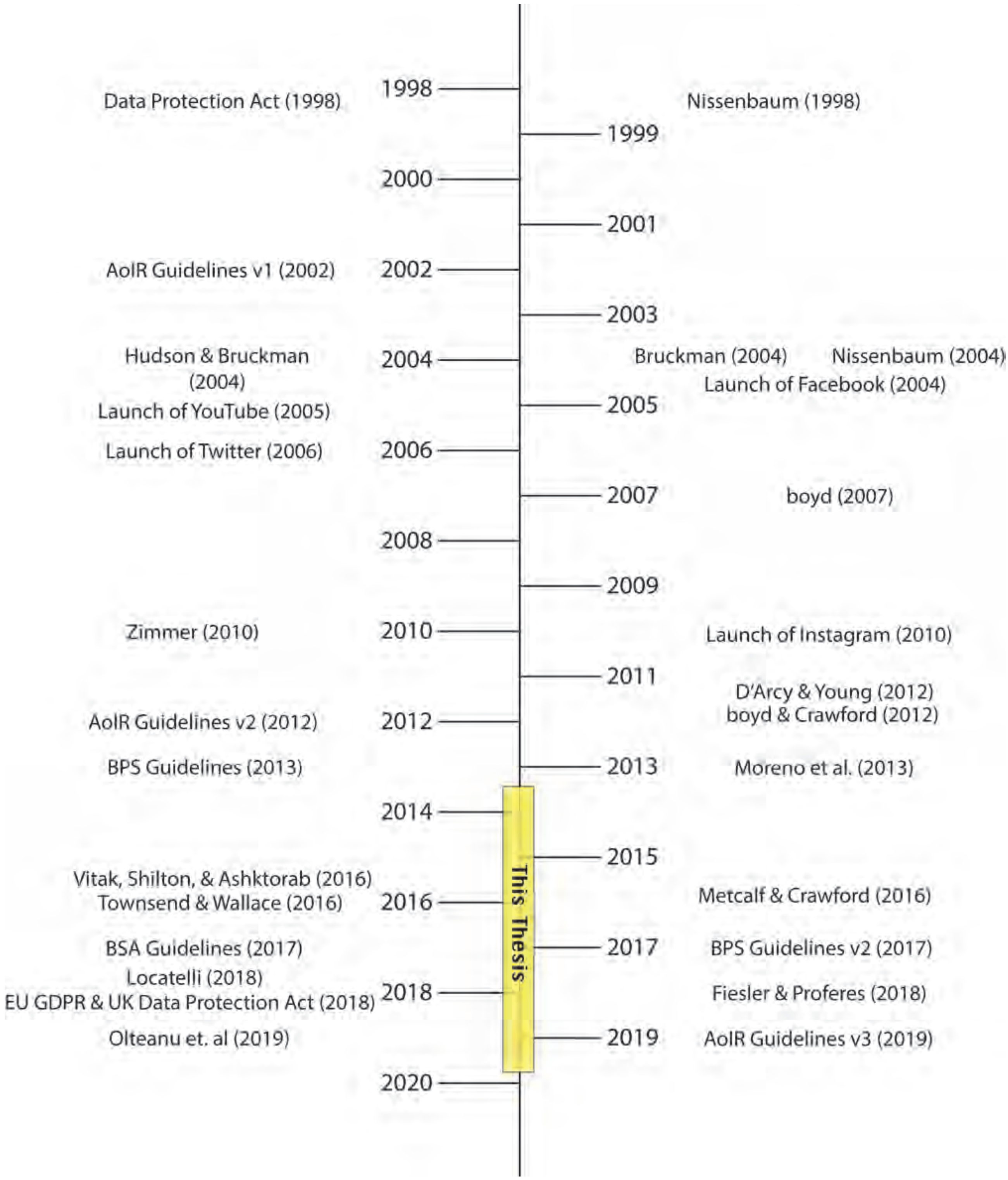


Figure 4.1: Timeline demonstrating timing of this thesis compared to legal, ethical, and academic best practice publications.

4.6.5.1 Informed Consent

Exactly how, and to what extent, informed consent can and should be gained from individual social media users, was and continues to be, a point of debate in ethical guidelines and academic literature. While the consensus, it could be argued, is that gaining the informed consent of thousands, if not millions, of social media users is impractical, and therefore not required [8, 32], there is also a shared view that where the intention for the content to be public is clear, that informed consent is not required [191].

In order to reduce and mitigate any potential issues around informed consent, the decision was made to primarily focus on content created and shared from the public social media accounts of identified organisations. Further to this, the information collected and stored that related to the social media accounts of individual users was kept to a minimum, with no personal information collected. This allowed for the relevant guidelines to be adhered to, whilst also reducing the risk of harm to any individual social media users who may be included as part of the collected datasets.

4.6.5.2 Maintaining Anonymity

Maintaining the anonymity of participants is a key aspect of the ethical guidelines and frameworks that have been described in previous sections. The process of anonymising data is designed to protect individual research participants and any information that may be considered personal or private, thus satisfying legal requirements, such as the Data Protection Act (1998) [1], and more recently the 2018 Data Protection Act [2] and EU GDPR [63]. As with with many aspects of ethical approaches to online and social media research, whether this can be achieved completely, and in all circumstances, is still a point for debate.

Whilst specific ethical considerations in relation to particular social media platforms are outlined in the following sections, the broad consensus between academics and many ethical guidelines is that ‘conversations’ (in this case, tweets and Facebook posts) should not be repeated verbatim in publications. Search functionality on social media platforms, or indeed search engines, could facilitate the finding of the original author by searching for the provided quote. As such, as a general rule, any quotes from individuals’ accounts, used as examples or to demonstrate a point are paraphrased, or merged together. In doing so, the original context and meaning can be preserved, while making it more difficult for the original authors to be identified. Further to this, any accounts belonging to individuals that are mentioned in the tweets or Facebook posts are replaced with generic terms, such as @user, for example.

In the studies presented here, only social media content created and shared from the social media accounts of organisations was collected and analysed, and as such anonymity in this context did not need to be maintained. However, it was still possible that such posts may include mentions of the accounts of individuals. In such contexts, then appropriate measures would have been taken, such as replacing these mentions with @user, for example. Further steps have been taken when reporting aspects of the social network analyses, such as not indicating exactly which user accounts are critically positioned within the network, for example.

4.6.5.3 Profiling Individual Users' Behaviour

One of the main objectives of the research presented in this thesis is to develop a technique for profiling, or understanding, the engagement behaviours of social media users. The notion of profiling individual users, and acting on this information, could be interpreted by some as being ethically dubious. Indeed, social media platforms often have policies in place that prevent (or at least have strong repercussions for those that flout the policy) automated behaviour based on keywords or phrases. For example, Twitter will block applications (i.e. 'bot' accounts) that automatically respond to accounts that include a given word or phrase in a tweet.

The idea of being profiled, on an individual basis, is also likely to raise questions from members of the public, and would perhaps require more consideration of the notion of informed consent (as discussed previously). Further, tailoring social media content for targeted delivery to specific individuals is also likely to result in drastically increased overheads - both in terms of time and finances - for organisations.

However, the methods proposed here (particularly Chapter 7), while considering engagement on an individual basis, aggregates this data, thus mitigating this risk. Further to this, the focus is on supporting the development of social media strategies that encourage the engagement of a wide subset of followers, rather than targeted delivery direct to specified individuals. This approach has the added benefits of reducing the likelihood of individuals feeling 'targeted' by the organisation, and aiding the organisation in their social media content strategies. The details of this approach were included in the relevant ethical approval applications, which are described in Section 4.6.6.

4.6.5.4 Platform-Specific Ethical Considerations

Within the studies presented in this thesis, data was collected from both Facebook and Twitter. The ethical considerations surrounding the collection and use of data from these platforms in this context is discussed in the following sections.

Facebook Data

In the Facebook-based study (see Chapter 5), only posts sent from organisations were collected and analysed. Given the nature of these organisations, and the fact that the Facebook pages are dedicated public-facing pages for those organisations, it was determined that informed consent was not required - the previously discussed guidelines and frameworks support this as there was no "*expectation of privacy*" [8, 32, 148], nor was there any personal information being disclosed. The only data collected that may relate to private individuals was the number of likes per post. However, given that only the number of likes was collected, rather than the details of the specific individuals liking the posts, this again reinforced the decision that informed consent was not required, as no individuals are identifiable from this collected data.

Twitter Data

The studies presented in Chapters 6, 7, and 8 use data collected from Twitter. The Twitter API (at the time of writing) provides access to data that could be considered as identifiable, and as such presents more extensive ethical considerations. The data available through the API which was used in these studies can be categorised as follows, each category is then discussed in turn:

- Information relating to retail locations' accounts and the posts sent from these accounts.
- Information made available by users as part of their profiles - names, location information, and biographical information.
- Information associated with accounts, such as who the accounts follow, and who follows the accounts.
- Information which indicates how the user has engaged with specific social media content - i.e. retweets, likes, etc.

Twitter Data - Organisations' Tweets. For the studies in Chapters 7 and 8, it was necessary to collect the tweets sent from various organisations' Twitter accounts, within a specified time frame.

If these accounts belonged to individual members of the public, then there may have been a need to distinguish between what may be classed as 'public' or 'private' data (as described previously). However, as these accounts belong to the retail locations, and are used as part of their digital promotion, it can be assumed that each post is deliberately shared and intended to be public - that is, that there is no "*expectation of privacy*" [148].

As such, most, if not all, of the collected tweets could be reported verbatim (as examples of particular types of content, for instance), with the only notable exception being tweets where accounts belonging to individual members of the public are directly mentioned. In these instances, it is common practice to replace the username being mentioned with a generic term, such as @user.

Twitter Data - Profile Information. In each of the studies which used Twitter data, it was possible to gain access to, and collect, information relating to the Twitter profiles of individuals. It should be noted, however, that accounts marked as 'private' are not accessible without the express permission of the account holder, and as such form no part of the datasets used within this thesis.

There is an extensive list of data fields that can be accessed for each public profile - including name, username, location (if the user has specified one), and the biography or user description field. Not all of this information was necessary for the studies presented here, and was not collected or stored, however, such information could be used to supplement the techniques and findings discussed within this thesis, and at such a point the ethics of handling such data would become increasingly relevant.

While some of this data may have been collected (such as user ID, follower numbers, etc), none of this data is directly reported within this thesis, or associated publications, on an individual basis, in relation to any results or created profiles. The collection and use of user ID numbers is necessary throughout the studies, in order to maintain a means by which a single account can be identified throughout the various data collection and analysis processes.

This data is freely available through both the Twitter API (excluding 'private' accounts), as well as through the Twitter web and app interfaces. As the user should be aware of what information they are providing, and that it is clearly visible on their public profiles, it could be argued that this satisfies the 'information is public' criteria, as suggested by Moreno et al [140]. As with other Twitter-based data, such information is not available for accounts marked as being private, or deleted, and as such form no part of the datasets used within this thesis.

Twitter Data - Relationship Data. In the studies presented in Chapters 6, 7, and 8, the collected data includes relationship links between accounts - namely, those accounts that follow each other. This data is available through the public API (in machine-readable format), and also available through the web and mobile applications. As such, this can be considered as being public information, as it satisfies the conditions suggested by Moreno et al [140] as outlined earlier. Such data is not available from accounts marked as 'private', and as such, any private accounts and their relationships with other accounts is not included in the data used within these studies.

Twitter Data - Engagement Information. Engagement information (i.e. which tweets a user has retweeted) was collected for the studies discussed in Chapters 7 and 8. As with the previously discussed categories of Twitter data, this data is publicly available through the API, and web and mobile applications. As such this again meets the 'published publicly' criteria outlined by Moreno et al. [140].

4.6.6 Ethical Approval Processes

In this section, the various ethical approval processes that were undertaken during this research are outlined. Due to moving between institutions, changes in institutional approval processes and systems, and the period of time covered by the research included in this thesis, the inclusion of all relevant paperwork has not been possible - copies of ethics applications that could be sourced are included in Appendix A. However, the approval processes undertaken at the University of Lincoln are detailed below, followed by a discussion of those undertaken at Northumbria University.

4.6.6.1 Ethical Approval Process - University of Lincoln

Ethical approval was sought at the beginning of this research process, at the University of Lincoln, from the ethics committee with oversight over the School of Computer Science. This application laid out the general goals of the overall research, as well as the specifics of the planned studies - including the data to be collected, and the planned analyses. Citing the Data Protection Act (1998) as the relevant legal framework at that time, and both the ethical guidelines of both the Association of Internet Researchers, and the British Psychological Society, the application for ethical clearance addressed the three main issues discussed in previous sections: informed consent, maintaining anonymity where relevant, and the ethical concerns around 'profiling' individual social media users. Further, this application also made clear the plans for minimising the amount of data collected to only that which was necessary for the purposes of this research; storing the data securely, and retaining the data for only as long as is necessary.

Due to the nature of the data being collected, and the steps described previously, this research was judged to be low-risk, and was approved, with further ethical clearance required if the planned research was drastically changed.

This ethical clearance covered the period in which all of the data was collected, from both Facebook and Twitter, as well as the processing of all of this data for the purposes of the studies presented within this thesis.

4.6.6.2 Ethical Approval Process - Northumbria University

As a result of moving PhD registration from the University of Lincoln to Northumbria University, it was necessary to summarise the ethical clearance provided to date, as well as to outline planned further data collection and analysis, in order to confirm that the research could continue. This transfer took place in October 2016, and as such was prior to the introduction of GDPR (and Data Protection Act, 2018), as well as the updated ethical guidelines provided by the British Psychological Society (2017), or those released by the British Sociological Association (2017). As a result of this, the legal frameworks and professional guidelines referred to in the initial application at the University of Lincoln were still relevant.

At this point, ethical clearance was sought by January 2017, with the data processing already completed, with the remaining planned activities focusing on the interpretation of these analyses, production of publications and the writing of the thesis itself.

4.7 Overview of Studies and Methods Used

Study one, detailed in Chapter 5, serves as an initial investigation into the use of social media by retail locations, including the types of content that are produced and shared through these social media accounts. In doing so, this study addresses the first objective, and contributes to addressing the second objective. The findings of this study help to justify, contextualise, and inform, the remaining studies.

Study two, presented in Chapter 6, looks to build on the findings of the first study by tracking the development of networks of social media followers over time. Further, social network analysis techniques are applied to this data to assess the validity of such metrics in predicting online follower 'churn'. The findings of this study address both the second and third objectives, and also inform the fourth study.

Study three, which presents a method for generating engagement profiles for social media users, forms Chapter 7. This study builds on the findings from study one, which suggested that individuals' own preferences may affect their engagement behaviour. In doing so, this study addresses objective four, and further informs the fourth study.

Study four draws on each of the previous three studies, and combines user engagement behaviours with more traditional social network analysis techniques. In doing so, the role of such behaviours in predicting user retention, and by extension, the growth of online communities, is explored. This study is detailed in Chapter 8.

The methods used in each of these studies are now discussed in the following sections.

4.7.1 Study One - Understanding Retail Locations' Use of Social Media

This introductory study aimed to develop an understanding of how individual retail locations make use of social media, and in doing so, provide a context and grounding for the studies that follow.

4.7.1.1 Data Collection

Choice of Platform and Retail Locations
Facebook was selected as a data source for this study, making use of Facebook's public API [64]. Facebook is a commonly adopted platform for organisations such as retail locations, and provided suitable methods for collecting data covering a prolonged period of time, as well as the level of engagement each of these posts received. At the time of data collection, posts shared from public Facebook pages could be collected, along with associated data and meta-data, such as public comments, and the number of times the post was liked or shared. As a result of the access limitations at the time of collection, it was not possible to collect any data that users would have marked as being private, or from accounts with more restrictive privacy settings.

Since the data for this study was collected, Facebook has further limited access to such data, focusing on providing more 'meaningful' access to organisations with what it considers to be genuine business and operational reasons for accessing and making use of such data. While this may limit the access that is available to individuals and researchers (in some cases) it would not limit the access to the same, or equivalent, data for organisations, or companies designing and implementing applications for such organisations. As such, while the original means of accessing this data may no longer be available, any results and implications for use by organisations are still valid, and could be implemented by those with access to such data. Wider implications around issues such as limited access to data, are described later in this thesis.

For this particular study, six retail locations were selected to act as a case study. These locations were selected as they each maintained an active presence on Facebook, are based in a range of geographical areas within the United Kingdom, and operate independently of each other. As such, these should represent a range of commercial organisations, with differing target markets, and as such, differing online audiences.

Data Overview
The Facebook posts authored and sent from six distinct Facebook accounts were collected, covering a 12-month period, from 1st October 2012 to 30th September 2013. These six locations are: Bluewater Shopping Centre (Kent), Boxpark Shoreditch (London), Brent Cross Shopping Centre (London), Bullring & Grand Central Shopping Centre (Birmingham), Oxford Street (London), and Trinity Leeds (Leeds). A brief overview of each location is provided in Chapter 5 (Section 5.2), with a fuller description of the locations and their online presence provided in Appendix B.

Using the Facebook API, each public post sent from those accounts was collected, along with the number of likes that each post received. This led to a dataset comprised of 3,759 posts and 135,513 associated likes. Further dataset details can be seen in Table 5.1.

4.7.1.2 Qualitative Analysis of Post Content

As discussed in previous sections of this chapter, reflexive thematic analysis was deemed to be an appropriate approach to the analysis of qualitative content - in this case, Facebook posts. In this study, all six stages of the process outlined by Braun and Clarke were followed: data familiarisation, generating initial codes, searching for themes, reviewing themes, defining & naming themes, and, finally, reporting (as described previously, in Section 4.3.2). Due to the quantity of social media posts being analysed across each of the accounts, this process was time consuming, a point which is reflected on later in this thesis.

4.7.1.3 Calculating Engagement Rates

One aspect of this study involved considering the levels of engagement achieved through the posts shared from the retail locations' Facebook accounts.

As outlined previously, and detailed more fully in Chapter 5, the total number of likes for each post was recorded - a total of 135,513 likes. There are multiple ways in which this data could be handled and analysed: total number of likes, mean number of likes per post, and mean likes per post as a percentage of total followers. Each of these approaches are outlined below.

Total Number of Likes

This method is the most simple way of assessing the overall engagement on a collection of social media posts and is formalised in Equation 4.1. By totalling the number of likes (in this case), it would be easy to rank the retail locations, and their approach to social media, based on the total number of likes that they received. However, this would not take into consideration the number of posts that were created and posted by each location. For example, an organisation that shared one post would be measured, in all likelihood to their detriment, in the same way as an organisation that posted one hundred times.

$$Total\,Likes = Likes_1 + Likes_2 + ... + Likes_i + ... + Likes_n \tag{4.1}$$

Mean Number of Likes per Post

This approach (shown in Equation 4.2) builds on the previous approach, but takes into account the number of posts sent from that account within the time-frame being analysed. As such, it negates the potential 'advantage' that would be given to organisations that post significantly more than others. By dividing the total number of likes (as above) by the total number of posts sent, organisations' posting frequency is negated, giving a more even view of engagement gained from social media posts. However, this may still favour those organisations with an inordinately higher number of followers. For example, an organisation receiving 90 likes from ana audience of 100 would be 'disadvantaged' when compared to an organisation receiving 180 likes from an audience of 10,000.

$$Mean\,Likes\,Per\,Post = \frac{Likes_1 + Likes_2 + ... + Likes_i + ... + Likes_n}{Total\,Number\,of\,Posts} \tag{4.2}$$

Mean Number of Likes per Post, as a Percentage of Account Followers
This approach (demonstrated in Equation 4.3) takes into account both the number of posts shared by the account during the period of analysis, and the total number of account followers at that point in time. This negates any potential 'advantage' that may be experienced as a result of either posting much more than other accounts, or having a much larger audience than other accounts.

$$Mean\,Likes\,Per\,Post\,as\,a\,Percentage\,of\,Followers = \left(\frac{Mean\,Likes\,per\,Post}{Total\,Number\,of\,Account\,Followers}\right) \times 100 \tag{4.3}$$

Implementation in this Study
In this study, 'Mean Number of Likes per Post, as a Percentage of Account Followers' was selected as the most accurate and useful measure for engagement. Measuring this engagement on a month-by-month basis allows for this to be compared to the type of content being produced and shared each month, to identify any potential correlation between the types of content being shared and the level of engagement achieved as a result. The results of this are discussed in detail in Chapter 5. Although only considering Facebook 'Likes' in this particular study, this method of calculating and presenting engagement is similar to methods widely used by organisations and platforms offering social media analysis products to organisations and individuals.

Investigating Engagement Rates over Time
In this study, the relationship between engagement rates each month, and the type of content shared that month was also investigated. In order to assess this relationship, a series of Spearman's rank correlation coefficients (a non-parametric test) was selected as an appropriate means of determining the strength of relationship between content type and engagement levels, for each retail location. As a non-parametric test, no assumptions need to be made about the distribution of the data; Spearman's rank focuses on the relative ranks of the two variables being investigated. As with other statistical tests conducted within this thesis, p values for each of the results are considered, to indicate the significance of the results.

4.7.2 Study Two - Social Media Follower Churn

The aim of this study was to analyse the development of online follower communities, in a longitudinal study, and determine the feasibility of using social network analysis techniques and metrics to predict those users that are likely to 'churn' (see the Glossary for more detail) and leave the network.

4.7.2.1 Data Collection

Choice of Platform and Retail Locations
The main focus of this study is the growth and development of follower communities that follow retail locations on social media. Due to the type of data required - namely lists of followers, and the relationships between them, Twitter was selected as the most suitable social media platform for this particular study.

In order to collect the required data, the Twitter API was used, in particular the 'Followers' and 'Friends' endpoints [194], which return the lists of the accounts that follow, and are followed by, the specified account.

For this study, six retail locations were selected to act as case studies. These locations were selected as they cover a range of geographical areas, are active on social media - particularly Twitter, and would have a range of different target demographics and markets.

Use of Data 'Snapshots'

In order to track the evolution of social graphs over time, a '*snapshot*' method was adopted. In this instance, a snapshot is defined as one complete instance of social graph data, based on a single initial account. These include each public follower of the account, the relationship between these followers, as well as their own followers who are not (yet) following the main account being analysed. Snapshots were collected, for all six locations, on a continual basis over an 18-month period. The six retail locations selected were: Regent Street, London; Carnaby Street, London; Seven Dials, London; Covent Garden, London; and London West End. A description of these retail locations is provided in Appendix B.

Data Overview

Data was collected for six retail locations, over an 18-month period. Each dataset consisted of the followers of a retail location, and the accounts that they follow, and are followed by. This allows for the relationships between retail accounts' followers to be demonstrated, but also allows for 'peripheral' users to be identified - those that follow, or are followed by, followers of the retail account, but who do not yet follow the retail locations' account themselves.

4.7.2.2 Construction of Social Graphs

Data collected from the Twitter API includes lists of users that followed the particular retail locations' social media accounts, along with the users that followed, and were followed by, these followers. From this data, it was possible to construct a list of edges that demonstrate the relationship between each pair of accounts, based on whether A follows B, or vice versa.

While there are many applications that will handle and process such data in order to form social graphs, or networks, such as Gephi [14], many of these can experience limitations when handling large datasets. 'R' [163] and associated applications such as 'R Studio' allow for the programmatic import, handling, and processing of data (thus increasing repeatability of this approach) and can handle much larger datasets. For these reasons R, R Studio, and associated packages, were selected as an appropriate environment for handling and processing the data for this aspect of this study.

4.7.2.3 Selection and Use of Social Network Analysis Metrics to Test Hypotheses

The hypotheses posited in this study were developed from existing work in this field, and related fields, of research, and relate to the application of 'traditional' social network metrics in modelling and predicting the growth or decline of these networks.

Each hypothesis, detailed fully in Chapter 6, is tested by calculating the specified metric for each snapshot, and measuring its impact in predicting those who leave the network between each snapshot, or over a series of snapshots. In each instance, a null hypothesis positing no relation between the relevant variables is tested, with the alternative hypothesis then accepted if the null hypothesis is rejected. This is done using logistic regression models, using the metric as an independent variable, and the outcome (i.e. whether the user remains as part of the network, or leaves the network) as a dependent variable.

Logistic regression techniques are often used in the field of churn prediction [167] and are seen as suitable in this context as they allow for the modelling of multiple predictor variables against a given outcome variable. The results generated from these regression tests are multi-faceted; relevant elements have been outlined previously in Section 4.4, and are further discussed in Chapter 6.

Statistical Significance

As discussed previously, in Section 4.4.4, an appropriate level of significance is required before a null hypothesis can be rejected. This level of significance is referred to as the alpha level. In this study, an alpha value of .05 was set. Where the indicated p value for each test is less than or equal to this alpha level, the null hypothesis can be rejected, and indicates a statistically significant result. Where the p value is above this level, then the test has failed to reject the null hypothesis, and the result has been indicated to be statistically non-significant.

4.7.3 Study Three - Modelling User Engagement with Social Media Content

In this study, a method for generating profiles for users' engagement with different types of social media content was demonstrated and evaluated. This also involved grouping similar engagement profiles together, and comparing these groupings to network communities determined using social network analysis techniques.

4.7.3.1 Data Collection

Choice of Platform and Retail Locations

The aim of this study was to build on the findings reported in study two and develop a flexible technique for creating engagement profiles for social media users. Twitter was selected as an appropriate data source for this study, as it not only acted as a natural follow-up to study two, but also provided data relating to a useful form of engagement - the retweet.

In order to collect the data necessary for this study, the Twitter API was used [194] - in particular, the 'User Timeline' endpoint, which returns the tweets sent from the specified account. At the time of collection, access to public data was available to individuals through unique sets of API keys, which allowed the rate of access to data to be restricted. Other than this, collecting public tweets sent from specific accounts was essentially unrestricted. As well as basic data, such as the content of a tweet, the account from which it was posted, and when it was posted, various meta-data fields were also made available (although not collected or used within this study) [194].

Since data collection for this study was completed, access to Twitter APIs, and as a result the data offered through the API, has been restricted further. Developers and researchers must now go through a more rigorous application process to gain access to the API, which is now restricted to those with 'valid use-cases', including business use-cases. This is still somewhat less restricted than current Facebook access restrictions (which requires valid business use-cases and applications for each individual API endpoint); once a use-case has been approved then users can access data from all public accounts.

Data Overview

Two retail locations - intu Metrocentre (Gateshead) and Regent Street (London), were selected to act as case studies. These locations were selected as they are geographically separate, but both represent locations with large number of tenant retailers, and are active on social media - particularly Twitter. An overview of these locations is provided in Chapter 7 (Section 7.2), with a fuller description provided in Appendix B.

The gathered data consisted of tweets sent from these accounts over an 18-month period, as well as social graph data over the same time period. A total of 6,784 tweets were collected, along with 17,943 users as part of the two social graphs. A full description of this dataset can be found in Chapter 7 (see Section 7.2).

4.7.3.2 Qualitative Analysis of Post Content

This study involved the qualitative analysis of Twitter content, in a similar manner to the Facebook post content analysed in the previous study.

Following the same process, as described previously in Section 4.3.2.1, and as detailed by Braun and Clarke [27, 28, 29, 30], both the main dataset and validation dataset were qualitatively analysed using reflexive thematic analysis. While reflexive thematic analysis generally follows six defined stages, this process omitted some of the stages - namely the identification of themes - in order to construct the engagement profiles (this process is described in detail in the next section). This was to allow the construction of engagement profiles based on the granular data of individual codes, rather than broader themes. However, such a process could be repeated and based on the broader themes.

4.7.3.3 Constructing Engagement Profiles

With the collected data analysed, and the content codes checked and refined where necessary, it was then possible to create engagement profiles for each user, based on the tweets with which they had engaged.

These engagement profiles are represented as n-dimensional vectors (with n equal to the number of content codes created and applied to the dataset), with each value (or dimension) in the vector representing the percentage of engaged-with tweets that contained that particular content type. For example, user who engaged with a total of 3 tweets, with one engaged-with tweet coded as 'Product Comments' would have a value of 0.33 in the 'Product Comments' element of the vector. The process of creating these engagement profiles is detailed more fully, using data from this study, in Chapter 7.

Creating profiles in this way allows for this approach to be used with any content type, and dataset size. Content codes can be easily applied to text, still images, video content, and audio, and the size of the vectors will reflect the amount of different codes that have been created. This allows for the creation of techniques and processes that are flexible in nature, and can be used in a variety of organisational contexts - a need that has been highlighted in related work discussed previously.

4.7.3.4 Clustering Engagement Profiles

Rather than focus on individual engagement profiles at this stage, users were segmented into groups containing users with similar engagement profiles. These groupings were determined using k-means clustering - a widely used clustering algorithm that is an efficient and low-cost means of clustering multivariate data [100]. This clustering algorithm requires the value of k to be predetermined, one suitable method of doing so (and the method used within this study) is through the use of the gap statistic [189]. Further, the Euclidean distance was selected as the distance metric to be used within the clustering algorithm - this is widely used as it is deemed to produce favourable results when used with data with a high number of dimensions [3]. As profiles created using this method could (in theory) contain any number of dimensions, using a distance metric that is scalable is crucial, and again this allows for the flexible use of this approach in a range of contexts.

With the clusters of similar engagement profiles created, it is then possible to determine the *general engagement behaviours* of each cluster, which were determined using the centre points of each cluster. This highlights dimensions (and thus context types) that were present (or indeed, absent) in a high percentage of the engaged-with tweets. These *general engagement behaviours* show the overall engagement preferences around which the cluster was formed. These behaviours (in this particular study) fell into 8 different categories, although this will of course differ, based entirely on the context and data on which the profiles are based. These behaviours, and the overall process are explained in greater detail in Chapter 7 (see Section 7.5).

4.7.3.5 Generating Social Graph Communities

Determining communities within social graphs [144] is common practice, with related work described previously in Chapter 2. As a well-documented, explored, and supported field of work, existing packages [51] within the R software application [163] facilitate this. Based on the collected data which had been formed into social graphs, the graph for both retail locations was grouped into communities of highly-connected users. This was done using the modularity metric, which is commonly used in identifying communities within social graphs [19], based on the strength of ties. The specific results of this are detailed more fully in Chapter 7 (see Section 7.6) with details provided as to the number of communities, the size of these, and how they compare to the clusters of engagement profiles that have been generated previously.

4.7.3.6 Comparing Social Graph Communities and Engagement Clusters

One aspect of this study was to compare how groups of users, clustered around common engagement behaviours, might compare to communities of users based on their positions and relationships within a social graph. The methods by which these two sets of groupings were achieved are explained in previous sections of this chapter.

This comparison was made using two different methods. The first involved taking each engagement cluster in turn, and calculating the percentage of its members that were assigned to each of the identified social network communities. The second adopted a similar approach, but considered the members of each social network community, and calculated the percentage of its members that are assigned to each engagement cluster. This process, and the accompanying results are demonstrated in detail in Chapter 7 (see Section 7.6).

4.7.4 Study Four - User Engagement and Social Graph Growth

In this study, the impact of engagement on predicting and modelling user churn is investigated, building on the findings of previous studies. This is done through combining the results of the previous studies, with the number of observed engagements with the analysed social media content.

4.7.4.1 Data Collection

Data used in this study included the dataset used previously in the study presented in Chapter 6, as well as the number of times each user present in that dataset engaged with the social media content in each of the 'snapshots' that were collected. As such, no new data was collected for use within this particular study.

4.7.4.2 Hypothesis Testing

For this study, each of the hypotheses was evaluated through the use of logistic regression, in a similar way to study two, presented in Chapter 6. Again, the same metrics were used to determine the validity of the models using these engagement figures, as well as comparisons with the models used in study two. In doing so, it can then be demonstrated if evidence of prior engagement with the account is likely to contribute to the user remaining as part of the online network following that account. As in previous studies, the alpha level for rejecting the null hypothesis was set at 0.05.

4.8 Summary

Building on the related works outlined in the previous chapters, this chapter has outlined the methodological decisions made, and approaches undertaken within this thesis. Following a discussion of various qualitative and quantitative approaches, the most appropriate approaches to the studies presented here were outlined. Following a discussion of relevant legal frameworks, ethical guidelines, and academic literature related to ethical concerns, each study has been outlined.

Table 4.3: Summary of a sample of 'good practice' and ethics-related academic literature.

Source Details	**Relevant Points Raised, Issues Highlighted, & Main Findings**
1998	
Protecting Privacy in an Information Age. (Nissenbaum) [147]	- Highlights the debate over the use of the terms public and private, particularly in relation to online activity. - As people are essentially willingly sharing information that is going to be analysed, they are, in some sense, contributing to the violation of their own privacy. - Technology and networking will facilitate large scale collection and storage of data.
2004	
Opportunities & Challenges in Methodology and Ethics. (Bruckman) [42]	- Human subject research norms do not apply to material that is published - the nature of online content makes it more difficult to accurately distinguish between what is considered 'published' and that which is not.
"Go Away": Participant Objections to Being Studied and the Ethics of Chatroom Research. (Hudson & Bruckman) [97]	- Chatrooms are often seen as 'private' spaces by participants. - Internet researchers olten publish in multidisciplinary journals. As such, some form of minimal ethical guidelines are needed "*so that these different professions can appropriately talk to one another*".
Privacy as Contextual Integrity. (Nissenbaum) [148]	- Advances in storing, aggregation, analysis, and mining of information, both online and offline leads to more questions regarding the notion of privacy in these contexts. - Cites two criteria for 'a reasonable expectation of privacy': 1) that the person exhibited an actual expectation of privacy, and 2) the expectation is one that society recognises and reasonable.
2007	
Why Youth (Heart) Social Media. (boyd) [24]	- Defining the term 'public' is difficult at best, and there are multiple 'publics'. - "*In reference to actions or texts, public often implies that the audience is unknown and that strangers may bear witness.*" [24, p.125] - There are 4 fundamental properties that distinguish unmediated publics from networked publics: persistence, searchability, replicability, and invisible audiences.
2010	

Table 4.3: Summary of a sample of 'good practice' and ethics-related academic literature.

Source Details	**Relevant Points Raised, Issues Highlighted, & Main Findings**
"But the data is already public": On the Ethics of Research in Facebook. (Zimmer) [217]	- Uses the T3 (Tastes, Ties, and Time) dataset released in 2008, as a case study. - Despite claims that "*all identifying information was deleted or encoded*", the identity of the source of the dataset was quickly discovered. - The re-identification of the source institution demonstrates how fragile the presumed privacy of data subjects included in such datasets. - Ways in which this dataset violated privacy: amount of personal information collected, improper access to personal information, unauthorised secondary use, and errors in personal information. - The researchers failed to "*respect the expectations likely held by the subjects regarding the relative accessibility and purpose of their Facebook profile information*". - "*future researchers must gain a better understanding of the contextual nature of privacy in these spheres.*" [217, p.323]
2012	
Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. (boyd & Crawford) [25]	- "*Big data is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets*". - All researchers are interpreters of data. - "*...it is problematic for researchers to justify their actions as ethical simply because the data are accessible...The process of evaluating the research ethics cannot be ignored simply because the data are seemingly public.*" - Biases and limitations must be acknowledged and understood. "*Without those biases and limitations being understood and outlined, misinterpretation is the result*". - "*the public discourse around such research [social media research] tends to focus on the raw number of tweets available.*" - Twitter studies are not representative of all people - Twitter users are a very particular sub-set of the wider population. Studies should be noted as such. - "*While some users post content frequently through Twitter, others participate as 'listeners'.*" - "*The size of data should fit the research question being asked; in some cases, small is best.*"

Table 4.3: Summary of a sample of 'good practice' and ethics-related academic literature.

Source Details	Relevant Points Raised, Issues Highlighted, & Main Findings
Ethics and Social Media: Implications for Sociolinguistics in the Networked Public. (D'Arcy & Young) [52]	- There are clear problems, challenges and limitations when conducting research online: "*The ramifications of this are acutely realised when human research ethics meet online social media*" - "*gives rise to unique ethical challenges*" - "*if the content is considered public, then the ways in which context is constructed and maintained through participation...must be considered*" - "*From an ethics perspective, this entails that individuals targeted for research in line social media are human subjects (Hudson + Bruckman, 2004, p. 128); their content incurs the same rights and obligations as does the content of offline engagement*" - Regarding Facebook - when collecting user data (i.e. private data, such as wall posts) using an application, users must be told what is being collected and for what aim. "*Users are thus afforded control over those aspects of their content which are not classified as public information on the site*" - There is still a distinction between posts made deliberately public (i.e comments on a public page), and private, i.e. wall posts etc. "*At the heart of the issue are fundamental misunderstandings concerning the contextual nature of privacy in online social media*"
2013	
Ethics of Social Media Research: Common Concerns and Practical Considerations. (Moreno, Goniu, Moreno, & Diekema) [140]	- Collecting and using social media data may not require informed consent from post authors if: • Access to the social media platform, and the post in question, is public • Whilst the information provided is identifiable, it is not considered private • Information gathering requires no direct interaction with the person who posted it online
2016	
Social Media Research - A Guide to Ethics. (Townsend & Wallace) [191]	- The context of the online setting itself helping to determine the applicability of informed consent. E.g. a post made within a protected, 'private' Facebook group would be considered private, and thus require informed consent from the author, whereas an open discussion on Twitter could be considered as being published publicly.

Table 4.3: Summary of a sample of 'good practice' and ethics-related academic literature.

Source Details	**Relevant Points Raised, Issues Highlighted, & Main Findings**
Where are Human Subjects in Big Data Research? The Emerging Ethics Divide. (Metcalf & Crawford) [133]	-"*There are growing discontinuities between the research practices of data science and established tools of research ethics regulation.*" - Relevant ethical frameworks for 'big data' studies are highly contested and in flux. - "*...there is a growing divide between established systems of research ethics in more traditional disciplines and the dynamic norms and research methods of Big Data.*" - Such research "*moves ethical inquiry away from traditional harms such as physical pain or a shortened lifespan to less tangible concepts such as information privacy impact and data discrimination.*" - There is still a real need to define what is a 'human subject' in such research studies, and more critically reflect on what is owed to 'data subjects' in such contexts. - "*Critical data studies have routinely demonstrated that it is deeply mistaken to treat research data as neutral and raw*" - "*...research ethics regulations are an imperfect codification of the hard-won, often-contested and evolving social trust invested in practitioners and researchers.*" - Such research can be seen as an oxymoron - when research is simultaneously considered both powerful and insightful about human lives, but inconsequential when considering potential harms. - While individual 'public' datasets may be considered inconsequential and relatively 'low-risk', combining multiple such datasets together can result in radically different consequences, thus rendering the usual anonymization techniques and safeguards insufficient. - "*...we should reject the belief that the risk borne by research subjects depends on what kind of data is obtained and how, rather than what is done with the data.*".

Table 4.3: Summary of a sample of ‘good practice’ and ethics-related academic literature.

Source Details	Relevant Points Raised, Issues Highlighted, & Main Findings
Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community (Vitak, Shilton, & Ashktorab) [197]	- “*...researchers still struggle to balance research ethics considerations with the use of online datasets.*” - A continuing major challenge for online research ethics is creating ethical standards for disparate fields, each with their own backgrounds, methods and approaches. - Various studies have demonstrated that anonymised datasets can be de-anonymised when combined with other datasets. - “*Almost all researchers who cite the public nature of online data also said they try to anonymise or avoid identifying individuals in public datasets, signalling that even ‘public’ data is still seen as sensitive by most.*” - Gaining informed consent is not a reasonable expectation or requirement for all open online datasets. However, wherever possible, researchers should respect the norms of the contexts in which the data is generated. - “*The challenges of online data collection are extraordinarily nuanced, and reflect problems based on the difficulties of defining contexts and norms in online spaces.*”
2018	

Table 4.3: Summary of a sample of 'good practice' and ethics-related academic literature.

Source Details	**Relevant Points Raised, Issues Highlighted, & Main Findings**
"Participant" Perceptions of Twitter Research Ethics. (Fiesler & Proferes) [67]	- It is rare that those whose data is being studied are involved in the creation and development of ethical guidelines. - The Internet introduced additional complications to research ethics, including anonymity and confidentiality; informed consent; terms of service; relationships with participants, and dissemination of findings to participants; and the (re)definition of public spaces and public data. - "*...many IRBs consider the study of public data to not be under their purview*". - "*...prior work suggests users are broadly unaware of how the content they create flows within a large informational ecosystem that Twitter supports*". - "*This finding suggests that many Twitter users incorrectly believe that researchers cannot use tweets at all or must ask permission from the users, and a majority of that group believe that this is an ethical rule for researchers.*". - Consider removing or anonymising identifiable information when quoting tweets. - Quoting tweets verbatim without good reason is not recommended. - Publication of user identity should only occur when the benefits of doing so clearly outweigh the potential harms, or with permission of the user(s) involved. - Findings of this study suggest that users believe in privacy by obscurity. - Context in analysis and quoting of results is key - a study about a sensitive topic such as a medical condition, or drug use, would be a less appropriate context for quoting tweets than, for example, a study regarding television viewing habits.
Ethics of Social Media Research: Srate of the Debate and Future Challenges (Locatelli) [123]	- Summarises many previous works already noted here. - Due to the nature of social networking platforms, the dominant issue is not about revealing characteristics of specific individuals, but rather about sorting people into groups. - With anonymization of data there is a trade-off between how comprehensively the data is anonymised, and the level of quality of research outputs and findings.
2019	

Table 4.3: Summary of a sample of ‘good practice’ and ethics-related academic literature.

Source Details	Relevant Points Raised, Issues Highlighted, & Main Findings
Social Data: Biases, Methodological Pitfalls and Ethical Boundaries. (Olteanu, Castillo, Diaz, and Kiciman) [151]	- “*Biases in social data and the algorithmic tools used to handle it can have, as a result, far-reaching impact.*” - Overlooking inherent biases within the data collection, analysis, and reporting stages can (inadvertently) to discrimination. - “*User engagement with content and other users is influenced by algorithms that determine what information is shown, when it is shown, and how it is shown.*” - Though not always regarded as such (at least, historically), research on social data is human-subjects research. - Both scientists and journalists have, and continue to, press for greater scrutiny over the use of social media data, focusing on potential ethical pitfalls such as breaches of privacy, and further enabling racial, socioeconomic, or gender-based profiling activities. - Obtaining informed consent from millions of users is impractical. - “*The acceptance of terms of use may not fulfil the criteria of informed consent, as the often vague language alluding to ‘research use’ does not involve a disclosure of the specific elements relevant to a particular research program.* - Archiving personal data for too long a time period, and sharing poorly anonymised datasets can contribute to privacy breaches, as datasets can be combined to gain further insights about people without their knowledge.

Chapter 5

Investigating Retail Locations' Use of Social Media

5.1 Introduction

Organisations, including retail locations such as shopping malls and 'destinations' such as shopping streets, continue to maintain and develop their online and social media presence. In some cases, organisations are dedicating considerable resources - both in terms of personnel and finances - to maintaining a successful online presence.

In this chapter, a study is presented which sought to begin to understand how retail locations utilise social media, specifically Facebook, in terms of the types of content they produce and share through the platform. Throughout this study, it became apparent that retail locations utilise social media in different ways to each other, and as a result of this seemingly gain differing levels of engagement from their audiences, particularly as audiences develop numerically. Retention of customers, or potential customers, is noted as being of importance to retailers, and by extension, retail locations [115]. This engagement with social media content may play an important part in retaining these online followers [174], it therefore follows that developing an understanding of the role of social media content, the resulting engagement, and this effect on user retention, or churn, is of importance to those tasked with maintaining the online presence of retail locations. Each of these is explored in more detail throughout this thesis.

The remainder of this chapter is organised as follows. First a description of the data collected for use within this study is presented. The following section briefly describes the methods used during the study, as well as the results of the analyses. Finally, a summary of the main conclusions of this study, and implications for future related studies is provided.

5.2 Data Collection

For this study, six UK-based retail locations were selected for analysis. These locations were selected for a variety of reasons. First, each maintained an active presence on Facebook for the time period being analysed (1st October 2012 - 30th September 2013). Second, they should attract different audiences due to their locations and the fact that they offer differing varieties of retailers and experiences. The six retail locations used for this particular study were:

- Bluewater Shopping Centre (Kent) - an out-of-town shopping centre approximately 18 miles from the centre of London. Bluewater opened in 1999, and is currently home to approximately 330 stores, services, and vendors. [20].

- Boxpark Shoreditch (London) - a food and retail shopping mall made from shipping containers. Described by its creator as the 'world's first pop-up mall'. Boxpark opened in Shoreditch in 2011, and two other sites have opened since then [22].

- Brent Cross Shopping Centre (London) - Opened in 1976, Brent Cross is in North-West London, and features 120 retailers [31].

- Bullring & Grand Central Shopping Centre (Birmingham) - the current Bullring & Grand Central Shopping Centre opened in September 2003, and currently houses 140 stores, services and vendors. In 2004, it was the busiest shopping centre in the United Kingdom, with a reported 36.5 million visitors [43].

- Oxford Street (London) - a well-known shopping destination in central London, Oxford Street is over 1 mile long, and contains over 300 retailers [154].

- Trinity Leeds (Leeds) - opened in March 2013, creating over 3,000 jobs. Trinity Leeds currently contains 120 retailers and services [192].

A fuller description of each of these locations is provided in Appendix B.

Table 5.1: Number of posts collected from each location's Facebook account, and total number of likes.

Location	Posts	Post Likes
Bluewater	500	13,770
Boxpark	1,286	21,143
Brent Cross	673	8,709
Bullring & Grand Central	685	45,614
Oxford Street	408	13,918

Each Facebook post authored and posted from the official accounts of these six locations was collected from the official Facebook API [64], along with the number of likes that each post attracted; the number of likes being used as an indication of the level of engagement for each post. The number of posts, and total number of likes for each Facebook account are detailed in Table 5.1. The resultant dataset consisted of 3,759 Facebook posts. Likes were selected as an appropriate indication of engagement as each Facebook account can only 'like' a post once (thus can be used as indication of the number of accounts that have engaged). Likes also offer a lower threshold to engagement than, for example, commenting on a post.

5.3 Analysis and Results

Following the data collection, which was detailed in the previous section, this section describes the analyses conducted on the data, and the results produced from these analyses. This includes both the content created and shared by the retail locations, and the results user engagement with this content. A broader discussion of the implications of these results is provided in the next section of this chapter.

5.3.1 Investigating Content Shared by Retail Locations

A qualitative analysis of each of the 3,759 Facebook posts was conducted, using the techniques detailed by Braun and Clarke [27, 28, 29, 30] and discussed in more detail in Chapter 4. This analysis enabled a detailed understanding of the content of each of the Facebook posts that were shared, and makes it possible to identify commonalities and differences between each of the accounts, as well as changes within the accounts over time.

Portrayal as Corporation	Information Distribution	Sharing a Link
		Event Information
		Sharing Retailer Information
		Offer Information
		Competition Information
		Product Information
		Weather Information
		Job Opportunity Information
		Retail Opportunities
		Behind the Scenes
		Travel Information
	Promotion	Promoting Retailers
		Promoting Products
		Promoting Social Media Use
Portrayal as People	Inquiry	Asking a Question - Product
		Asking a Question - Other
		Asking a Question - Event
		Asking a Question - Food
		Asking a Question - Retailer
		Asking a Question - Competition
		Asking a Question - External Events
		Asking a Question - Personal
		Asking a Question - Offers
		Asking a Question - Wather
		Asking a Question - Media
		Asking a Question - Social Media
		Asking a Question - Apps
	Opinion	Commenting on Products
		Commenting on World Events
		Commenting on Media
		Commenting on Current Activities
		Commenting on Event
		Commenting on Retailer
		Commenting on Food
		Commenting on Celebrity

Figure 5.1: Hierarchy of codes applied to Facebook posts. (L-R: Second-order themes, first-order themes, content codes.

The qualitative analysis resulted in the generation of a final set of 35 codes, which are listed in Figure 5.1. Each of these codes were applied to multiple Facebook posts, and in many cases multiple codes were applied to a single post. After reviewing the codes and their related Facebook posts, the 35 codes were organised into 4 first-order themes, which themselves were organised into 2 second-order themes. The relationships between the 35 codes and their first and second-order themes are demonstrated in Figure 5.1. The themes are described in the sub-sections that follow, with example posts used to illustrate each theme.

5.3.1.1 Description of Themes

Second-Order Theme: Portrayal as Corporation

Posts within this second-order theme act to portray the retail location as a corporate entity. In doing so, a professional body is portrayed, through the sharing of corporate information, as well as promoting retailers within the location, for example. Posts within this theme can be further categorised into the first-order themes '*Information Distribution*', and '*Promotion*'.

First-Order Theme: Information Distribution

This theme relates to posts that are distributing useful information to those viewing the Facebook page, or the page's followers. This information may include information about promotional events or job opportunities, as well as specific retailers or offers. Posts such as *"Forever21 are holding a Student Night at Bluewater on October 3rd."* are included within this theme.

First-Order Theme: Promotion

Posts within this theme actively promote a retailer, product, or social media account. Within this theme, posts can be seen as 'directives', often explicitly telling the reader to carry out a particular task or action. These posts include posts such as *"Get down to John Lewis to find everything you need."* and *"Share this post and comment with your most hilarious fish pun to enter"*.

Second-Order Theme: Portrayal as People

Within this second-order theme, posts can be perceived as portraying the personal side of the retail location, through taking on conversational tones, and sharing opinions. Posts within this theme can be further categorised into the first-order themes '*Inquiry*' and '*Opinion*'.

First-Order Theme: Inquiry

This theme relates to posts, or elements of posts, that are asking questions of the audience. Common variations of this include asking for opinions on a given product, event, or retailer. For example, *"What do you think of the return of dungarees?"*. Within this theme, there are also posts that appear to take on a more personal tone with the audience, asking questions such as *"Did your parents used to write your name in the back of your clothes?"*. Posts within this theme are likely to be intended to encourage dialogue between the audience and the retail locations [112].

First-Order Theme: Opinions

Posts within this theme include those where the author, or the retail location, are seen to be sharing 'their' opinion on a given topic, usually a retailer or a product. These posts are likely to have been designed to promote the retailer or product, but are portrayed as the sharing of opinion, rather than explicit promotion of a retailer or product. Examples of posts within this theme include: *"We love these styles from River Island"*, and *"We're loving the Autumn/Winter collections in our stores."*.

5.3.1.2 Differences between Retail Locations

Each of the four identified themes: '*information distribution*', '*promotion*', '*inquiry*', and '*opinions*' were used by each of the six locations to varying degrees. While each location had posts that fell within each of these themes, the prevalence of these themes differs from location to location and over time, which potentially signals different strategies being taken by the retail locations. Based on the prevalence of these themes over the time period being analysed, four different strategies or patterns of content were identified, with each retail location falling into one of these strategies each month:

- Very strong focus on ‘information distribution’.
- Strong focus on ‘information distribution’ with elements of ‘promotion’.
- Strong focus on ‘information distribution’ with elements of ‘inquiry’.
- Strong focus on ‘information distribution’ with the remaining focus equally split between ‘promotion’, ‘inquiry’, and ‘opinions’.

This variety in approaches shows that retail locations are utilising social media in different ways to disseminate information to their audience. This could indicate conscious decisions to approach the use of social media in a particular manner - the development and use of social media strategies. Each of these approaches attracts different levels of audience engagement, which is detailed and explored later. In the following paragraphs, the differing approaches of each location are briefly outlined.

Figure 5.2 shows the share of posts in each theme, as a percentage, per month, for each of the six retail locations. Each of these is now discussed, in turn, in the following subsections.

Bluewater

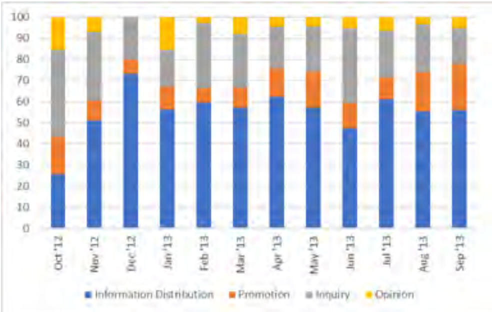
As demonstrated in Figure 5.2a, with the exception of October 2012, the dominant second-order theme within Bluewater’s posts was *Portrayal as Corporation*, which consists of *Information Distribution* and *Promotion*. Overall, there was a strong focus on *Information Distribution*, where, after October 2012, around 50% of posts each month could be considered as *Information Distribution*. The difference in posting behaviour that can be seen in October 2012 could be due to numerous events that were organised, each featuring high-profile guests, as demonstrated below:

> *Have we got any Olly Murs fans out there? In case you didn’t know (see what we did there?)...Essex’s favourite cheeky chap is coming to Bluewater on 29th October for his book signing.*

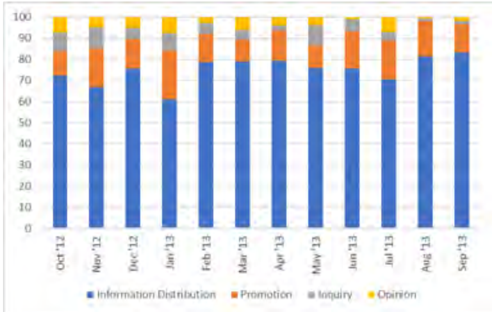
> *Please note that the time of the McFly book signing taking place this Thursday November 1st has been changed by Waterstones and will now commence at 11am and not 2pm as previously advised.*

> *This Saturday, 2008 Olympic Gold Medallist James Degale MBE will be defending his European Super Middleweight title at Glow, Bluewater. Tickets are selling fast but are still available, book yours now and don’t miss out on what’s sure to be a night of fantastic LIVE sporting action.*

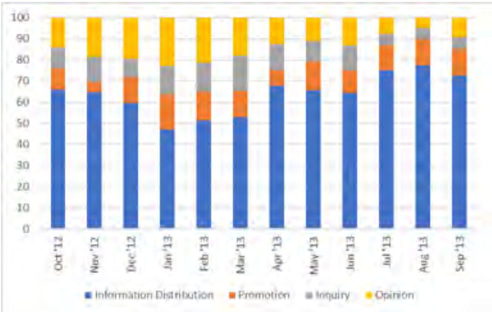
Overall, Bluewater appears to have maintained a relatively strong corporate appearance in terms of social media presence throughout the 12-month period being analysed. As discussed previously, the variations in behaviour can perhaps be attributed to significant events and developments within the retail location being represented.



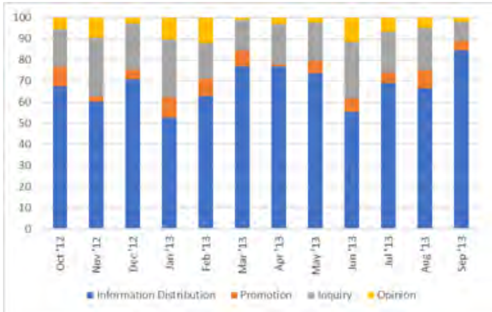
(a) Bluewater



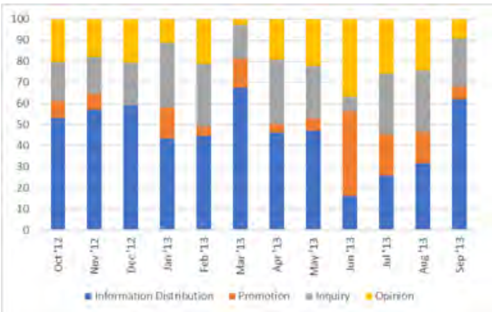
(b) Boxpark Shoreditch



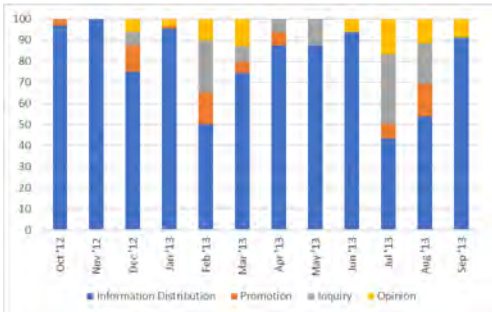
(c) Brent Cross



(d) Bullring & Grand Central



(e) Oxford Street



(f) Trinity Leeds

Figure 5.2: Share of posts in each theme, as a percentage, per month, for each retail location.

Boxpark Shoreditch

In a similar approach to that of Bluewater Shopping Centre, the majority of Boxpark's posts fall within the '*Portrayal as Corporation*' second-order theme (see Figure 5.2b), with over 80% of posts each month falling within this theme. As with Bluewater, the dominant first-order theme throughout the year was *Information Distribution*. The only real variation in behaviours was, perhaps, January 2013 where use of *Promotion* increased, and a minor increase was seen in both *Inquiry* and *Opinion*. Again, this can be attributed to temporary changes in the everyday running of the retail location, with a strong focus on January sales, as well as competitions that were being promoted at that time. Examples of these posts can be seen below.

> *Most caps (including these sick ones) at the New Era Europe Boxpark store: £10 EACH, 3 FOR 2. There's no possible way you can go wrong with that.*

Native Union SWITCH wireless, portable speaker is now in stock at Unit 10. Exclusive Boxpark launch price of £99 (RRP £130!)

Boxfresh have some crazy deals on in their sale. 70% off apparel. 50% off shoes. Spend £50 and get a selected item FREE.

Brent Cross

As with both Bluewater and Boxpark Shoreditch, *Information Distribution* is consistently the most prevalent theme each month (see Figure 5.2c). However, the content shared by Brent Cross contains more *Opinion* than the posts shared by Bluewater and Boxpark Shoreditch. This demonstrates that there is variation between the locations in terms of the types of content they are creating and sharing through social media. Examples of these posts include:

We are fighting over these amazing Reggae Reggae Groove Cut Crisps in our office.

We're loving the NEW Galton Flowers store.

In more general terms, Brent Cross is similar to both Bluewater and Boxpark in that, while the percentage of posts containing particular themes may vary from month to month, the ranking of these themes remains fairly consistent, i.e. after *Information distribution* the second most prevalent theme for Brent Cross is often *Opinion*. This suggests that there is some preference for the type of content that is produced for each account (what is unclear from this analysis is the reason for this) whether it is part of a formal social media strategy or not.

Bullring & Grand Central

Content shared from the Bullring & Grand Central Facebook account (see Figure 5.2d) follows the same trend as the previously discussed accounts - *Information Distribution* is the most prevalent content type, followed by a consistent second content type. In this particular case, *Inquiry* is the second most prevalent content type in each of the twelve months being analysed. Examples of these *Information Distribution* and *Inquiry* posts can be seen below.

This weekend we've celebrated our 10th birthday in style. Mollie King from The Saturdays presented The Show: Ten on Saturday 28th September and was then joined by her band mates for an exclusive performance on the catwalk.

Who's going to the Men's Health AW13 event at Superdry tonight?

In contrast to the previous accounts, there is only a minimum share of posts that contain either *Promotion* or *Opinion*-based posts. This reinforces the suggestion that, while the general behaviour patterns appear somewhat consistent across various Facebook accounts (i.e. majority of posts being information distribution), the finer detail of the types of content being shared differs between accounts.

Oxford Street

The Oxford Street account, in contrast to the other accounts described here, has a much more varied share of content types each month (as shown in Figure 5.2e). Whereas other locations predominantly produce *Information Distribution* content each month, the Oxford Street account behaviour varies, particularly in the summer months of June to August.

This change in content type is likely due to various events organised throughout the summer months. Posting behaviour also appears to have changed in December, though not in the same way as the content changes in the summer months. These two elements both suggest that change in focus of the retail location - in this case Christmas or summer events, are a contributing factor in changes in the types of content being created and shared by retail locations through social media. The two example posts below demonstrate the type of content that was being shared from the account over the Summer and Winter periods.

> *If you've got style and you're not afraid to show it then we want to see you on Oxford Street next week! We'll be out and about every lunch time spotting and rewarding the best dressed Oxford Street shoppers! Win anything from £5 - £50 worth of vouchers if we spot you looking stylish!*

> *Happy Christmas Eve all! Don't fret if you've still got presents to buy - shops are open 'til 7pm tonight! Ho Ho Ho..! :)*

Trinity Leeds

Of the six retail locations included in this study, Trinity Leeds is the only location which opened its doors to the public for the first time during the 12-month period being analysed - the location opened to the public in March 2013. As such, this presents an interesting opportunity to observe how posting behaviour may change in the transition from marketing a location that is yet to open, to marketing a location which is open to the public.

As with the other locations, *Information Distribution* remains the most prevalent theme throughout the 12-month period (see Figure 5.2f). In the first few months, posts are almost solely categorised as information distribution. This is quite possibly due to Trinity Leeds not yet being open for business. Information will have been shared relating to various retailers that would be opening within the location, as well as 'behind the scenes' information about the location as a whole.

> *We have added a number of career opportunities to our website including jobs at Carluccios, Fraser Hart, Giraffe, Mamas and Papas, New Look, Wagamama's and Fossil.*

The first major change in posting behaviour can be seen in February 2013 - the month prior to the location's opening. At this point, *Information Distribution* posts drop to around 50% of the content, with *Inquiry* and *Promotion* responsible for another 40% of content. Again, this suggests that retail locations adjust their social media content in response to major events or changes within the retail location. Other major deviations from *Information Distribution* can be seen in July and August 2013.

Again, this could be due to the location responding to changes and events - Krispy Kreme opened a store within the location in July 2013, and promotion for Students' Night (September 2013) would have taken place in August 2013.

> *We reached 10 million customers last week and to celebrate we pulled together this fun little video we hope you enjoy! A massive thank you to everyone who helped us pass this milestone! You're the best! And congrats Caroline, Nathan and family who scooped our "sur-prize"!*

As with many of the other retail locations, it appears that those responsible for the Trinity Leeds Facebook page have a natural tendency towards *Information Distribution* posts (either deliberately, on inadvertently), with this posting behaviour changing (albeit seemingly temporarily) due to situational changes - such as events.

5.3.2 Investigating Consumer Engagement with Post Content

Further to retail locations sharing various types of content, it also appears that their audiences respond to this content differently; this behaviour also changes over time, with content and engagement behaviours differing on a month-by-month basis. Figures 5.3 - 5.8 show the level of engagement per post, as a percentage of page followers for each location over the 12-month period. Presenting the data in this way normalises the level of engagement in relation to both the number of posts shared from the account, and the number of followers that each account had at that time. As such, it does not disadvantage those accounts that post fewer times than others, or those that have a smaller audience.

5.3.2.1 Differences in Engagement Patterns between Locations

Figures 5.3-5.8 demonstrates the levels of engagement, per post, as a percentage of the account's followers, for each location over the 12-month period. As can be seen from these figures, each location's engagement levels vary dramatically each month. For the majority of the analysed months, the levels of engagement received by each location are among similar orders of magnitude, with engagement often being less than 0.4%.

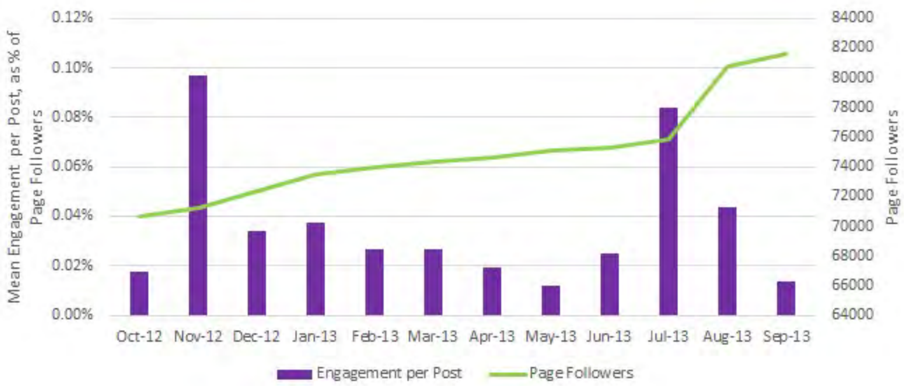


Figure 5.3: Average engagement per post, and page follower counts per month - Bluewater

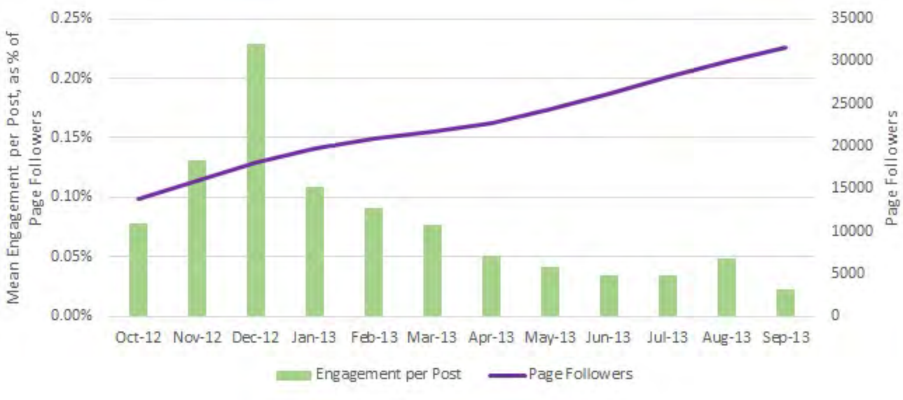


Figure 5.4: Average engagement per post, and page follower counts per month - Boxpark Shoreditch

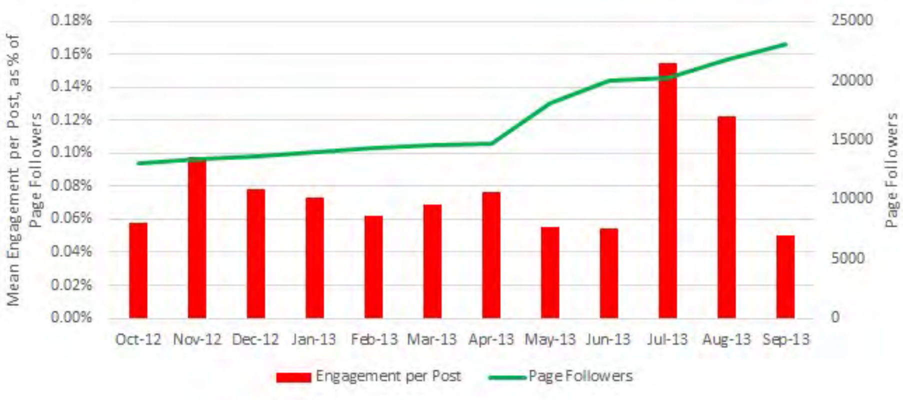


Figure 5.5: Average engagement per post, and page follower counts per month - Brent Cross

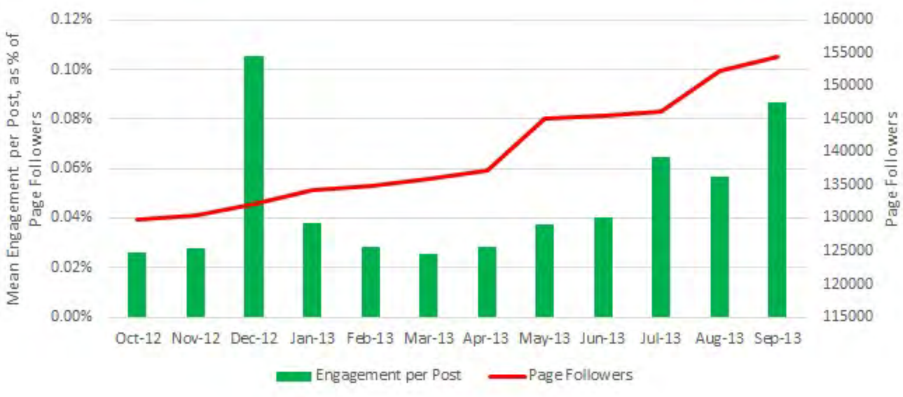


Figure 5.6: Average engagement per post, and page follower counts per month - Bullring & Grand Central

The main exception to this is Trinity Leeds (see Figure 5.8), which experienced much higher engagement levels prior to opening in March 2013. However, as can be seen from the figure, once the location opened to the public, the account followers increased dramatically, and the overall levels of engagement (as a percentage of total audience, per post) dropped considerably - to a similar order of magnitude as other locations. Boxpark Shoreditch (see Figure 5.4) experienced a general downward

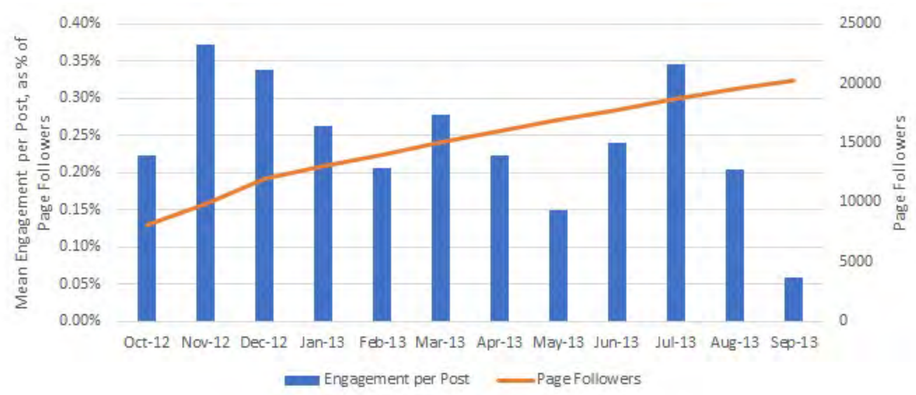


Figure 5.7: Average engagement per post, and page follower counts per month - Oxford Street

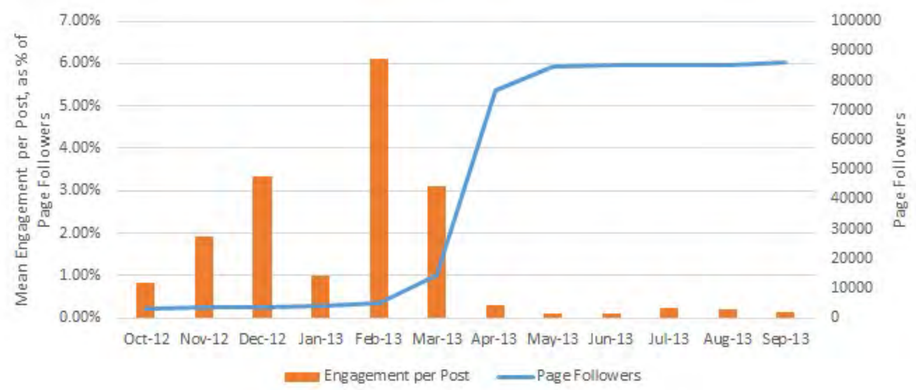


Figure 5.8: Average engagement per post, and page follower counts per month - Trinity Leeds

trend in engagement levels as follower numbers increased. Conversely, accounts such as Bullring & Grand Central (see Figure 5.6) experienced (in some months) increases in engagement as follower numbers increase. This would suggest that, perhaps, at least some of the new audience members may be more inclined to engage with content. It could also be explained as users beginning to follow the account after being exposed to content that they would readily engage with.

Though engagement levels are within similar orders of magnitude, there is no real consistency across the locations, even in terms of upward or downward trends as the size of audience increases. This leads to the suggestion that engagement, and engagement rates, are not directly connected to any one single factor, but rather are dependent on a combination of factors including both the audience, and the account being followed and engaged with.

Trinity Leeds (see Figure 5.8) provides a good example of what may happen to engagement levels as follower numbers increase dramatically - engagement levels dropped significantly. This change occurred as the retail location opened for business for the first time. Such a dramatic change suggests that audience makeup and motivations for following and engaging, before and after such a transition, are different. Those that followed the account before the location opened may have had different motivations for doing so, compared to those that began to follow the account later on. This suggests that audience makeup has an impact on the overall levels of engagement with the social media account. If this holds true, it also suggests that even in cases where the overall audience size remains the same - with individuals stopping following the account, and the same number of new followers beginning to follow the account - would also impact the engagement levels.

5.3.2.2 Comparison of Engagement Levels to Content Types

In the previous section of this chapter, the relationship between engagement levels and follower numbers was explored; in this section, potential links between social media content types and overall engagement is discussed.

Figure 5.2 shows the share of each content type, per month, for each of the analysed retail locations. Figures 5.3 - 5.8 show the levels of engagement for each location each month, as well as the changes in their follower numbers over the time period being analysed. When considering the relationship between content type and engagement levels, there are two main questions which will now be addressed:

- For each location, does the prevalence of one particular type of content correlate strongly with resultant levels of engagement?
- To what extent are these results consistent across all locations, in terms of which content types correlate with levels of engagement?

Table 5.2: Spearman's rank correlation coefficients, indicating strength of correlation between content type and engagement levels, for each retail location. Values in brackets indicate the p value associated with that result.

Retail Location	**Information Distribution**	**Promotion**	**Inquiry**	**Opinion**
Bluewater	0.070053 (.414)	-0.517480 (.042)	0.119089 (.356)	0.125874 (.348)
Boxpark Shoreditch	-0.419580 (.087)	0.097902 (.381)	0.461538 (.065)	0.440559 (.076)
Brent Cross	0.230769 (.235)	-0.269700 (.198)	-0.216780 (.249)	-0.069930 (.415)
Bullring & Grand Central	0.055944 (.431)	-0.174830 (.293)	0.181818 (.286)	-0.034970 (.457)
Oxford Street	0.000000 (.500)	0.181818 (.286)	-0.286710 (.183)	-0.062940 (.423)
Trinity Leeds	-0.129600 (.344)	0.466266 (.063)	0.083545 (.398)	0.121235 (.354)

In order to address the two questions posited above, Table 5.2 shows the Spearman's Rank correlation coefficients between the share of content themes and engagement levels each month, for each location. These correlation coefficients can range between 1 and -1, indicating perfect positive correlation and negative correlation respectively. A coefficient of 0 indicates no correlation at all.

Correlation between Content Type and Engagement for Individual Locations

As can be seen from Table 5.2, the majority of results are indicative of either 'very weak' (0.00 - 0.19), or 'weak' (0.20-0.39) positive or negative correlations between the types of content being shared, and the resultant levels of engagement.

Boxpark Shoreditch is the only account that, perhaps, has correlation results that indicate some kind of moderate relationship between most of the content categories and resultant engagement.

While *Information Distribution* posts demonstrate moderate negative correlation with engagement, *Inquiry* and *Opinion* posts demonstrate a positive moderate correlation.

There is no single retail location in this dataset that has any strong correlation between a type of content being shared and the resulting engagement, and very few of the results are deemed as statistically significant (p <.05). Indeed, three of the six retail locations demonstrate only 'weak' or 'very weak' correlations. This suggests that when considering an online audience as a single, homogeneous group, there is no single content or strategy that will encourage higher levels of engagement from the entire audience and as such, understanding the engagement preferences of an audience's constituent groups is of interest.

Similarities in Correlation across Retail Locations

When considering the correlation results across all six locations together, there is little in the way of pattern or commonality.

For example, *Information Distribution* has a moderate negative correlation when considering the Boxpark Shoreditch data, no correlation whatsoever when considering the Oxford Street data, and very weak positive correlation when considering the Bullring & Grand Central, and Bluewater datasets. Similarly, the prevalence of *Promotion* content has a moderate negative correlation for Bluewater, but a moderate positive correlation for Trinity Leeds.

Statistical Significance of Results

The p values for each of the results are indicated in Table 5.2. As can be seen, only one of these results is indicated as statistically significant (p <.050). Due to the number of tests conducted (24), this result also needs to be regarded with due caution, as there is a high probability (>70%) that such a result is a Type I error due to repeated testing.

Summary

This inconsistency between locations again highlights that each online audience is different, in terms of size, its constituent members, and their behaviours. These differences may then contribute to the variation in engagement behaviours from one audience to another. As such, adopting a 'one-size fits all' approach to social media content for retail locations is not valid. This further highlights the need to understand an audience, including their motivations and demonstrated behaviours, as was outlined in related literature in previous chapters.

5.4 Discussion of Findings and Implications

In this section, the main findings from this study are discussed. These include: the social media behaviours common across the accounts being studied; the distinct behaviours across these accounts; the dichotomy of professional and personable approaches to social media presence; and the changes and variation in audiences' engagement behaviours.

5.4.1 Common and Distinct Social Media Behaviours

The retail locations used within this study were selected as they appear to operate independently of one another, with each either maintaining their own online presence, or using a third-party to do so. The results, presented in Section 5.3 demonstrate that, for most retail locations, in most months, *Information Distribution* is the predominant theme of content being shared through social media. This is logical, given the affordances of platforms such as Facebook, which allows users - including organisations - to reach a wide audience, without geographical constraints [174].

This approach sees content being created that seeks to merely share information with the audience, rather than to facilitate 'relationship' between members of the public, and the organisation. Such an approach would also be expected if retail organisations, or those responsible for their social media presence, are approaching social media in the same way as they may approach more traditional media marketing.

Despite this common overall approach to creating social media content, the retail locations still approach social media use in their own unique way. From the number of posts shared each month, to the share of content type being created and shared, retail locations do not appear to follow a single, predetermined 'best-practice' approach to social media. This finding both begins to address, and further reinforces, the previously identified need to understand both what is being shared online [126], but also how individuals, and groups of individuals, are motivated to engage with this content, and why [171].

5.4.2 Organisations' Portrayal as Professional, or Personable

The dominant themes demonstrated for each location can often contrast against one another - the more traditional 'professional' view of marketing, which is sharing information and promoting, and the more 'personable' aspects afforded by social media - which aims to encourage engagement from and with the audience, and developing a relationship with them. Organisations are likely to have their own reasons for adopting either of these approaches, each of which have both advantages and disadvantages.

By remaining 'professional', and remaining (for the most part) monologic in their use of social media keeps a distance between the organisation and it's online social media audience. While this may not be conducive to developing relationships, it could be seen as a means to keep things simple, and merely using social media as an extension of their existing advertising strategies.

Strategies such as this can be planned ahead of time, with social media content developed as part of long and medium-term advertising strategies, which can then be checked and approved in the same way as other marketing campaigns. Maintaining this form of online presence may also be seen by those within organisations as a means of keeping control over exactly what content is shared, and when. Even seemingly well-planned social media campaigns can be received poorly by the public, and result in negative coverage which may have an impact on organisational performance. Adopting such an approach may be seen as a way of potentially mitigating these situations.

In contrast to this, some organisations appear to take a 'personable' approach to their social media presence. While this would still allow for some planning to be done, in terms of content to be posted, it encourages a more direct relationship between the organisation and it's audience, often resulting in dialogue between the two parties. Such an approach thus opens the organisation to needing to respond, in a timely manner, to the messages or comments posted by the public. In order to maintain this approach, it then becomes problematic to have each and every post and response planned or approved. As such, those operating the social media accounts have to be trusted to work independently, and within whatever guidelines have been set by the organisation (if indeed, they have been set).

The next challenge that may be encountered by organisations adopting this kind of approach, is deciding how to frame 'their' presence online: does the account represent the voice of the organisation, or is it clearly managed by a team of social media account managers? For example, the train company London North Eastern Railway (LNER) [122, 124], and Lidl - a retail store chain - often end posts with a caret symbol followed by the initials of the team member responsible for sending that post. This allows for individual members of the social media team to maintain their own 'personality' online, and thus develop their own rapport with the audience.

Research has suggested that dialogue enhances brand attitudes and purchase intentions, further suggesting that adopting this approach may have greater benefits for the organisations [48], although this also carries with it higher risks. One indication of this may be higher levels of engagement when organisations attempt to engage in dialogue with their audience. Again, this indicates that there is a continuing need to understand the motivations for individuals engaging with content online [60], and an understanding of how these motivations differ between individuals, and over time. This concept is explored and developed further in the following section.

5.4.3 Observed Variations in Engagement Behaviour

The results that have been outlined within this chapter demonstrate that there is no consistent level of engagement from an online audience (when considered as a single entity), either for one individual retail location, or when considering all six locations as a whole. This leads to the suggestion that, if organisations such as retail locations wish to tailor their online content to encourage higher levels of engagement, then they will need to consider the behaviours of individuals, or groups of individuals within the larger audience, rather than the audience as a single homogeneous group. This further reinforces the identified needs to not only understand what is being shared online [126], but also what motivates individuals to engage [60] and also how predictable this behaviour may be [171]. While these points are addressed to some extent in the chapters that follow, this is still a vast area of research that requires further attention.

Further to this, the results explored within this chapter have also highlighted that engagement behaviours from an audience appear to change and develop over time. While the data collected, analysed and explored within this chapter cannot be used to demonstrate exactly why this happens, it is suggested that this could be due to three potential causes:

1. Changes in the behaviour of an individual, or groups of individuals.
2. People that have engaged previously are leaving the 'audience' - i.e. stopping following the social media account.

3. New people that engage are joining the ‘audience’ - i.e. starting to follow the social media account.

Changes in the Behaviour of the Individual

It is possible that individuals who follow and engage with the social media account may change their behaviours over time. If this holds true, then the types of content with which they are likely to engage may also change and develop. Modelling, understanding, and (potentially) predicting this change will allow for content strategies to be adjusted accordingly. This will also begin to address the previously identified motivations for research, including the need to understand the predictability of individuals behaviour online [171], as well as understanding what is being shared [126]. Later chapters will explore this notion, considering how, and to what extent, the engagement behaviours of individuals develop and change over time.

People Leaving the Audience

The second suggested contributing factor to changing engagement levels is individual social media users removing themselves from the audience - i.e. stopping following a particular social media account. In doing so, they will reduce (or even eliminate) their exposure to the content being shared from the account, and thus reduce the likelihood of them engaging with the content. As previous research has suggested, this is likely to have an impact on the performance of the online presence of the organisation and, by extension, may impact the overall performance of the organisation.

While determining the exact cause for each individual acting in this way will be problematic, if not impossible, it may be possible for indirect actions to be taken that may encourage individuals to remain as followers of the social media account. This would require two main elements to be investigated; first, whether it is possible to predict which users are at risk of unfollowing, or leaving the direct audience; second, what types of content these individuals are likely to engage with. If this can be achieved, then it would be possible for social media account managers to identify those at risk of leaving, and then adjust their social media content to encourage them to engage, and remain as a follower of the account. This would not be possible for all accounts, of course, and the ‘value’ of retaining that particular user, or group of users, would have to be determined. If the value is too low, then it may be that no action is taken.

People Joining the Audience

As well as individuals leaving the network, new members joining the network may also impact the overall engagement with the social media account. Even when the total number of followers remain constant (i.e. the number of people leaving and joining balances out), this change in individuals is likely to change the overall engagement, and engagement preferences, of the audience. As such, understanding the engagement behaviours and preferences of an audience needs to be an ongoing process, even if the overall size of the audience appears to remain consistent.

5.5 Summary

This chapter has presented an initial study into the use of social media, specifically Facebook, by retail locations. Through an examination of the types of content shared by six separate retail locations, three main elements have been explored. First, the types of content shared by these locations over a 12-month period. Second, the numerical growth of these accounts in terms of the number of followers they have. Third, the average levels of engagement experienced by each of these accounts. The findings of this study, detailed in the previous sections, inform the design of the studies presented in the following chapters.

One finding from this study was the need to understand how, and when, individuals may leave, or join the network. That is, whether it is possible to predict with any degree of accuracy, which users are at risk of unfollowing a given account, and whether those loosely connected to the account may start to follow it. The study presented in Chapter 6 explores the notion of customer 'churn', and how this may be modelled and predicted through the use of social network analysis techniques and metrics.

Another finding of this study relates to understanding the engagement preferences of individuals, or groups of individuals, rather than treating the entire group of social media followers as a single homogeneous group. This finding both begins to address, but also highlights, the motivations from previously discussed literature, which identified an ongoing need to understand the motivations and sharing behaviours of individuals [60, 171]. As these motivations and preferences are based on complex individual choices and experiences, these engagement behaviours are likely to differ from individual to individual, and thus no audience will be made up of a group of individuals who all behave and engage in the same manner. This is explored further in Chapter 7, in which a method for profiling the engagement behaviours of individuals is proposed, implemented, and evaluated.

In the next chapter, Chapter 6, the development of online audiences, and the notion of follower 'churn' is explored, including the use of social network metrics as a means of modelling and predicting potential follower churn.

Chapter 6

Predicting Social Media Follower Churn Using Social Graph Metrics

6.1 Introduction

Many standard social graph metrics exist, and are used to indicate the state of a social graph at any single point in time. Often these are measured at varying intervals, in order to show how the graph is developing - be that growing, or declining. The study presented in the previous chapter demonstrates how retail locations make use of social media, and began to consider how types of content may affect the levels of engagement received on those posts. One element of the findings explored within the chapter suggests that the growth of a network can impact the levels of engagement received. Previous research discussed in Chapter 3 also suggests that any segmentation of an audience needs to stable for a reasonable period of time, to enable any insights based on that segmentation to be acted on. As such, the ability to model, and to some extent predict, how a follower network surrounding a given social media account may develop over time is likely to be of interest and importance to those managing an organisation's social media accounts.

The study presented in this chapter considers the use of various social network metrics as a means of predicting the development of a social network over time. Using multiple retail locations as case studies, and collecting 'snapshots' of the state of their follower networks over a prolonged period of time, the relevance of various metrics in predicting follower churn is analysed. This builds on the initial findings from Chapter 5, and lays the foundations for studies presented in later chapters of this thesis.

The remainder of this chapter is outlined as follows. First, various hypotheses are stated and described, based on the previously explored related literature. Following this, a brief overview of the collected and analysed data is provided. Third, the findings from these analyses are discussed, with each hypothesis addressed in turn. Following this, the application of these findings, both to studies presented later in this thesis, and potential further work is discussed.

6.2 Hypotheses

Based on the related research and literature explored earlier, the following hypotheses will be tested. A description of each, including the metrics that will be used to test these hypothesis is provided below. In each instance, a null hypothesis of no demonstrable relationship between the variables will be tested directly, with the following hypotheses forming the alternative hypothesis in each case.

H1. *Users who are more central within the network are more likely to remain as part of the network.*
This hypothesis is broken down into four sub-hypotheses. Each test the same overall concept, but are based on different SNA metrics. Though users are certainly not aware of these values and how they relate to their position within the network, they can be used as a proxy for how connected into the network they may be, or feel. Each of the selected metrics are calculated in different ways, and as such each is being used to test this hypothesis.
These hypotheses will be tested using the 'static' values for each snapshot - that is, the values as they were calculated for each snapshot, with the outcome being whether the account is still present within the network at the next snapshot.

(a) *Users who have a higher closeness centrality are more likely to remain as part of the network.*

(b) *Users who have a higher betweenness centrality are more likely to remain as part of the network.*

(c) *Users who have a higher transitivity measure are more likely to remain as part of the network.*

(d) *Users who have a higher eigenvector centrality measure are more likely to remain as part of the network.*

H2. *Users are more likely to remain in the network, if they also have followers who are part of the network.*
Previous literature has indicated that social media users have some sense of their 'imagined audience' [128], and manage their online behaviour accordingly. It follows that, perhaps, users also manage who they follow, and engage with, in order to maintain a stable online presence, and thus remain relevant to their own followers.
This hypothesis will be tested using a boolean yes/no value, which indicates whether or not the user has followers who are also part of the network, and the dependent variable, indicating if they are still present in the following snapshot of the network's development.

H3. *The more followers a user has in the network, the more likely the user is to remain a part of that network.*
This hypothesis is related to hypothesis 2, but considers the effect of the number of followers on the churn behaviour. Extending the previous hypothesis, this assumes that a user is more likely to be conscious of their audience, and adjust their behaviour accordingly, if the number of relevant followers is higher.

H4. *Users are more likely to remain in the network if they also follow other people who are part of the network.*
This hypothesis is related to the concept of homophily [54, 131]; if the user follows accounts that also follow the main account, it suggests that there may be a greater level of shared interests, and as such may improve the likelihood of them remaining as part of the network.

H5. *The more accounts a user follows who are part of the network, the more likely the user is to remain a part of that network.*
This hypothesis builds on the previous hypothesis, again related to the concept of homophily [54, 131], which suggests that individuals seek out and connect with those that they perceive to be like themselves. If users are following many accounts that are also following the main account, this suggests there is an increased level of shared interests, and as such that the main account may be of more interest to them, hence them being more likely to remain as part of the network.

H6. *Users with higher amounts of followers are more likely to remain in the network.*
Many 'popular' accounts, i.e. those with many followers, may see their online following as a commodity, and thus be more considered in changes in their online behaviour that may jeopardise this following [128]. If so, changes such as unfollowing accounts (and thus potentially losing regular access to the content shared from these accounts) is likely to be a more considered decision, and may result in them remaining as part of the network of account followers.

H7. *Users who are becoming more central are more likely to remain as part of the network.*
These hypotheses, which measure changes in various centrality measures, are based on the premise that if a user is shown to be moving towards the centre of the network - thus being more connected - then they are unlikely to suddenly stop following the main account. If this holds true, then users that are seen to be slowly moving 'out' of the network would be seen to be at risk of churning.

(a) *Users whose closeness centrality value is increasing are more likely to remain as part of the network.*
(b) *Users whose betweenness centrality value is increasing are more likely to remain as part of the network.*
(c) *Users whose transitivity measure is increasing are more likely to remain as part of the network.*
(d) *Users whose eigenvector centrality measure is increasing are more likely to remain as part of the network.*

H8. *Users are more likely to remain as part of the network if their 'in-network' follower count is increasing.*
In a similar sense to the previous hypothesis, users that are seen to be increasing their number of followers that are also following the main account are, potentially, unlikely to cease following the main account. This is again related to the concept of homophily [54, 131], which suggests that people are more likely to connect with those that they perceive to be like themselves.

H9. *Users are more likely to remain as part of the network if the number of 'in-network' accounts they follow is increasing*.

This hypothesis is based on the supposition that if users are seen to be taking steps that are increasing their connections with other accounts that also follow the main account, then it is unlikely that they will suddenly cease to follow that account. If this holds true, then users that are reducing this number of connections may be deemed to be at risk of churning.

H10. *Users are more likely to remain as part of the network if their number of followers is increasing*.

An increasing number of social media followers could be used to demonstrate an increased, or sustained, popularity. This popularity, in part, may stem from the information that is shared (or perhaps, retweeted) through the user's account. This may lead to a more considered approach in terms of who the account follows, thus reducing their likelihood of churning.

H11. *A combination of metrics will be more effective at predicting when a follower may leave, or remain in, the network, than any single metric*.

While each of the previous hypotheses tests a single metric in relation to predicting social media follower churn, this hypothesis posits that a combination of these metrics will be more effective in predicting and modelling follower churn.

6.3 Data Collection and Analysis

For this study, six retail locations were selected to act as case studies. These locations include: Regent Street, London; Carnaby London; Seven Dials, London; Oxford Street, London; Covent Garden, London; and London West End. A detailed overview of these locations can be found in Appendix B. Data collection processes were conducted over a 16-month period, from October 2014 through to January 2016. This study focuses on six retail locations all within the London area, thus acting - to some extent - in direct competition to one another. With retail areas within London being a competitive market, online audience retention for each of these organisations will be of utmost importance.

For each of the stated hypotheses, the null hypothesis (h_0) - that there is no statistically significant relationship - will be tested. If the significance of the results is above the specified α level ($>.05$) then the null hypothesis is accepted. Where the statistical significance of the results is below the specified α level, then the null hypothesis is rejected, and the alternative hypothesis can be accepted.

In the first group of hypotheses (H1 - H10), each hypothesis was tested individually, and consisted of a single metric, and an indicative outcome - either that they remained a follower in (or left) the network. Depending on the hypothesis being tested, the data being used is referred to as either 'static' or 'dynamic'. The static datasets include SNA metrics calculated for each node in the network for that particular snapshot. Dynamic datasets are based on the change between these values in-between each pair of snapshots.

Following this, the final hypothesis (H11) relates to a combination of metrics and whether, by using a combination of metrics, the potential for follower churn can be predicted more effectively than through the use of individual metrics.

To test each hypothesis, a logistic regression was used. Logistic regressions are often used as part of churn modelling, across a range of fields and contexts [167], and as such was deemed as a suitable method to utilise in this study. The results from each hypothesis' testing are described, in turn, in the following section of this chapter.

6.4 Results

This section includes a description of the results for each test, in order to address the hypotheses that were posited earlier in this chapter. For each hypothesis, a summary of the results will be provided, before being discussed in relation to the original hypothesis. The first step for each of the posited hypotheses was to test the associated null hypothesis, determining the statistical significance of any relationship between the variables being tested, before considering the effect size of any relevant relationships, and the goodness of fit of the model.

6.4.1 Hypothesis 1

As outlined previously, this hypothesis relates to the centrality of the user within the network being analysed, and this position in relation to the likelihood that the user remains a part of the network. This hypothesis is broken down into four sub-hypotheses, which test four separate centrality measures.

6.4.1.1 Hypothesis 1A

Hypothesis 1A is: "*Users who have a higher closeness centrality are more likely to remain as part of the network*"; as such, the null hypothesis being tested is "*There is no statistically significant relationship between closeness centrality and remaining as part of a network*". In order to test this hypothesis, the closeness centrality measure for each node, for each of the six retail locations was recorded, for each of the recorded snapshots. In addition, the presence of each node in the following snapshot was recorded, to use as the boolean result - if true, the user remained as part of the network, if false they had left the network before the next snapshot was recorded. For each snapshot, for each location, the statistical significance of the result was recorded; a summary of these results are presented in Table 6.1.

Table 6.1: Hypothesis 1A. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	Mean p Value
Regent Street	0	1	2	.340
Carnaby London	0	0	0	.480
Seven Dials	0	2	2	.370
Oxford Street	0	0	0	.482
Covent Garden	0	2	3	.303
London West End	1	2	2	.310
Total	1	7	9	.381

The results demonstrated in Table 6.1 show that only 4 of the relevant snapshots, across all six locations have results with a p value of $<.050$, with mean p values ranging between .303 and .482. With a low percentage of snapshots showing a significant relationship between the two variables, and the mean p value being so high suggests that there is no definitive, statistically significant, relationship between the two.

As such, in this case, h_0 - the null hypothesis - is accepted, there is no consistent statistically significant relationship between closeness centrality and a user remaining part of an online follower network in this case.

6.4.1.2 Hypothesis 1B

This hypothesis is similar to hypothesis 1A, but uses the betweenness centrality metric, as opposed to closeness centrality. In this instance, the null hypothesis being tested is "*There is no statistically significant relationship between betweenness centrality and remaining as part of a network*". Table 6.2 demonstrates the statistical significance of the relationship between the variables, as indicated by the logistic regressions that were conducted.

Table 6.2: Hypothesis 1B. Summary of statistical significance of results from logistic regression testing.

Location	$p<.001$	$p<.050$	$p<.100$	Mean p Value
Regent Street	0	0	0	.554
Carnaby London	0	1	1	.484
Seven Dials	0	0	0	.586
Oxford Street	0	1	1	.448
Covent Garden	0	1	2	.621
London West End	0	1	1	.483
Total	0	4	5	.529

In much the same way as hypothesis 1A, the number of snapshots which demonstrate a statistically significantly relationship between the tested variable and the user remaining part of the network is very low. Combined with a mean p value of .529 across all of the retail locations, this suggests that the null hypothesis is supported by the data and is therefore accepted, there is no statistically significant relationship between the two variables.

6.4.1.3 Hypothesis 1C

Again, similar to hypothesis 1A & 1B, this hypothesis tests a centrality measure as a means of indicating whether a user is likely to remain part of an online network. This hypothesis makes use of the transitivity metric. As such, the null hypothesis, h_0, in this instance is "*There is no statistically significant relationship between transitivity and remaining as part of a network*". The statistical significance of the results from the logistic regression tests can be seen in Table 6.3.

Based on the results presented in Table 6.3, the null hypothesis can be rejected, the results demonstrating that this particular metric could be used, to some extent, in order to predict if a user is likely

Table 6.3: Hypothesis 1C. Summary of statistical significance of results from logistic regression testing.

Location	$p<.001$	$p<.050$	$p<.100$	Mean p Value
Regent Street	1	4	4	.205
Carnaby London	2	7	8	.108
Seven Dials	0	2	3	.236
Oxford Street	1	4	5	.256
Covent Garden	1	4	4	.243
London West End	3	6	6	.134
Combined	8	27	30	.197

to remain part of the network. For some locations, such as Carnaby London and London West End, the majority of snapshots have statistically significant results.

With statistically significant results demonstrated, and the null hypothesis rejected, the next step is to determine the size and nature of the relationship between the variables. Full results tables, for each retail location, can be found in Appendix C, but a summary of these is provided in Table 6.4.

Table 6.4: Hypothesis 1C. Logistic regression results table.

Location	Estimate	Std.Error	Null Dev.	Residual Dev.	AIC	R^2
Regent Street	0.5565	0.3075	1943.2	1938.9	1942.9	0.003545
Carnaby London	0.6845	0.2621	2920.5	2910.9	2914.9	0.003565
Seven Dials	0.5605	0.4049	1450.8	1447.1	1451.1	0.003091
Oxford Street	0.2699	0.2533	2265.8	2256.1	2260.1	0.003381
Coven Garden	0.4462	0.2357	3767.6	3762.5	3777.6	0.001832
London West End	0.5194	0.2183	4658.9	4650.2	4654.4	0.001861
Overall	0.5062	0.2803	2834.5	2827.6	2833.5	0.002879

From the results summarised in Table 6.4, it can be seen that each retail location has a positive relation between the transitivity centrality value of a social media account, and the likelihood that they will remain a part of the network. That is, the more central a user is (using the transitivity measure), the more likely they are to remain following the account for the near future. Conversely, those with lower transitivity values are thus more susceptible to churning. However, the Nagelkerke's R^2 metric is low, suggesting that other models may be more effective.

6.4.1.4 Hypothesis 1D

This hypothesis tests the use of the eigenvector centrality metric as a means of predicting the likelihood of users remaining a part of the network for the near future. As such, the null hypothesis being tested can be posited as: "*There is no statistically significant relationship between eigenvector centrality and remaining as part of a network*".

Table 6.5: Hypothesis 1D. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	Mean p Value
Regent Street	0	2	3	.235
Carnaby London	0	1	2	.229
Seven Dials	0	1	1	.480
Oxford Street	1	4	4	.391
Covent Garden	1	8	8	.013
London West End	5	8	9	.013
Total	7	24	29	.364

Unlike the results of hypotheses 1A-1C, the results for this hypothesis, recorded in Table 6.5, show mixed results for the various retail locations. Covent Garden, for example, has statistically significant results for each snapshot, while both Carnaby London and Seven Dials have statistically significant results in only 11.11% of snapshots. The range of results suggests that the null hypothesis is tentatively rejected and the alternative hypothesis accepted and that the metric could be of use in predicting follower churn; results for each snapshot can be found in Appendix C, a summary of the results can be seen in Table 6.6.

Table 6.6: Hypothesis 1D. Logistic regression results table.

Location	Estimate	Std.Error	Null Dev.	Residual Dev.	AIC	R^2
Regent Street	5.6450	4.6913	1943.2	1939.8	1943.8	0.001921
Carnaby London	1.7583	4.4340	2920.5	2918.3	2922.3	0.000906
Seven Dials	1.2813	4.0473	1450.7	1449.5	1453.5	0.000876
Oxford Street	9.9498	5.6078	2265.8	2257.3	2261.3	0.003251
Covent Garden	17.2935	6.2308	3767.6	3757.6	3761.6	0.003291
London West End	29.6781	6.8478	4658.8	4622.6	4646.6	0.006878
Mean	10.9343	5.3098	2834.4	2824.2	2828.2	0.003358

As can be seen from Table 6.6, the effect size for this metric is much higher than the variable in hypothesis 1C (see Table 6.4). However, as the two variables being tested in these two hypotheses have different ranges, and typical values, this cannot be used as a straightforward comparison. For all but one of the retail locations, using eigenvector centrality as a predictor for individuals leaving the network is a better model than using transitivity (indicated by a lower AIC value), though both can still be used to some extent. The Nagelkerke's R^2 values indicate that this model still only explains a very low percentage of the overall outcome. Based on these results, it is accepted that hypothesis 1D is supported by the data. This would indicate that users with a higher eigenvector centrality value are less prone to churning, whilst those with a low value are more prone to churning.

6.4.1.5 Summary of Hypothesis 1

This hypothesis was broken down into 4 sub-hypotheses, each of which addressed using one centrality measure as a means of predicting whether or not individuals are likely to 'churn' and stop following the retail location's social media account. Of these four, two metrics were deemed to produce enough statistically significant results to be considered as viable for predicting (in a limited fashion) online social media follower churn. As such, this hypothesis is partially accepted - subject to the type of centrality measure being used.

6.4.2 Hypothesis 2

This hypothesis, as stated earlier, is "*users are more likely to remain in the network, if they also have followers who are part of the network*". Accordingly, the null hypothesis being tested is: "*There is no statistically significant relationship between users having followers who are part of the network, and the user remaining as part of a network*". In order to test this hypothesis, a binary variable - 'followers are in network' was used as the independent variable in the regression tests. A summary of the statistical significance of the results for each of the six retail locations can be seen in Table 6.7.

Table 6.7: Hypothesis 2. Summary of statistical significance of results from logistic regression testing.

Location	$p<.001$	$p<.050$	$p<.100$	Mean p Value
Regent Street	0	2	4	.237
Carnaby London	0	1	2	.428
Seven Dials	0	2	3	.449
Oxford Street	1	1	1	.410
Covent Garden	0	4	4	.350
London West End	1	4	4	.308
Total	2	14	18	.364

The significance results, summarised in Table 6.7, highlight that having followers that also follow the retail location rarely has a statistically significant link with remaining as a follower of the account, as such the null hypothesis is accepted; although each location has some statistically significant results, this metric cannot be used to reliably predict if a user will remain a part of the network in the near future.

6.4.3 Hypothesis 3

This alternative hypothesis: "*the more followers a user has in the network, the more likely the user is to remain a part of that network*" has a corresponding null hypothesis: "*there is no statistically significant relationship between the number of in-network followers a user has, and that user remaining as part of that network*". This is assessed using the number of 'in-network' followers the user has, in the logistic regression model. A summary of the significance of the results for each location, based on the collected snapshots can be seen in Table 6.8.

Table 6.8: Hypothesis 3. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	Mean p Value
Regent Street	0	0	0	.454
Carnaby London	0	0	0	.428
Seven Dials	0	2	2	.355
Oxford Street	0	0	1	.373
Covent Garden	0	0	0	.522
London West End	1	1	1	.379
Total	1	3	4	.419

As can be seen from the summary table, there are very few instances where the link between the number of 'in-network' followers, and the likelihood of remaining a follower of the account is statistically significant. For four of the six retail location accounts analysed, there was no snapshots in which the results are statistically significant at p<.050 and as such, in this case, the null hypothesis (h_0) is accepted.

6.4.4 Hypothesis 4

The stated hypothesis: "*users are more likely to remain in the network if they also follow other people who are part of the network*" is based on the concept of homophily and network ties [54, 131]. If users are also following 'in-network' accounts, this would suggest - perhaps - a strong affiliation between the accounts, be that common interests, demographics, or something other contributing factor. If this holds true, it would be expected that 'churning' in these particular accounts would be lower, and as such those who follow 'in-network' accounts are likely to remain a part of the network in the near future. Here, the null hypothesis being tested is: "*There is no statistically significant relationship between users following in-network accounts, and that user remaining as part of the network*". An overview of the statistical significance of the logistic regression results can be seen in Table 6.9.

These results demonstrate that, although there is no overall consistency between the six retail locations, this metric can be used (to some extent) in predicting whether a user is likely to remain as a follower of the account in the near future - for two of the locations, the results are statistically significant (at p<.050) in all of the data snapshots recorded. For only one of the locations was there no statistically significant results. As such, the null hypothesis is rejected, and the alternative hypothesis is accepted. A summary of effect sizes and associated data is provided in Table 6.10, full tables for

Table 6.9: Hypothesis 4. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	Mean p Value
Regent Street	1	4	4	.243
Carnaby London	1	4	4	.299
Seven Dials	0	0	1	.457
Oxford Street	1	1	2	.351
Covent Garden	8	9	9	<.001
London West End	8	9	9	<.001
Combined	19	27	29	.225

Table 6.10: Hypothesis 4. Logistic regression results table.

Location	Estimate	Std.Error	Null Dev.	Residual Dev.	AIC	R^2
Regent Street	0.3086	0.2624	1943.2	1938.8	1942.8	0.002886
Carnaby London	0.3964	0.2194	2870.5	2915.2	2919.2	0.002218
Seven Dials	-0.1104	0.4111	1450.8	1449.7	1453.7	0.000876
Oxford Street	0.3082	0.1939	2265.8	2244.9	2248.6	0.005251
Covent Garden	0.8469	0.1770	3767.6	3747.5	3751.5	0.006589
London West End	0.9602	0.1526	4658.8	4602.8	4606.8	0.011037
Mean	0.4637	0.2361	2826.1	2816.5	2820.4	0.004810

each retail location can be found in Appendix C.

As can be seen from Table 6.10, the effect of the variable varies between retail location, but in general, there is a positive relationship between following other 'in-network' accounts, and remaining part of the network. The AIC values suggest that, by a very small margin, a model using this metric is better than those models using centrality measures. Nagelkerke's R^2 values also indicate that this model is a slight improvement, though it does also indicate that there is still significant room for improvement.

6.4.5 Hypothesis 5

This hypothesis: "*the more accounts a user follows who are part of the network, the more likely the user is to remain a part of that network*" builds on hypothesis 4, and tests the relationship between following an increasing number of 'in-network' accounts, and the likelihood of remaining a part of that network. As such, the null hypothesis being tested, h_0, is: "*there is no statistically significant relationship between the number of 'in-network' accounts the user follows, and the user remaining as part of a network*". A summary of the statistical significance of the results from each snapshot of each retail location can be seen in Table 6.11. Though each retail location has some snapshots where this metric has a statistically significant relationship with retaining users as followers, there is a relatively low percentage of snapshots where this is the case, and the mean p values are high. As such, in this instance, the null hypothesis is accepted - there is no clear significant relationship and therefore use of this metric in predicting follower churn.

Table 6.11: Hypothesis 5. Summary of statistical significance of results from logistic regression testing.

Location	$p<.001$	$p<.050$	$p<.100$	Mean p Value
Regent Street	0	1	2	.463
Carnaby London	1	4	4	.351
Seven Dials	0	2	4	.320
Oxford Street	1	3	3	.266
Covent Garden	0	2	2	.344
London West End	0	3	3	.355
Total	2	15	18	.350

6.4.6 Hypothesis 6

This hypothesis (the alternative hypothesis) relates to the number of overall followers that an account has - rather than the number of followers that are 'in-network': "*users with higher amounts of followers are more likely to remain in the network*". Accordingly, the null hypothesis being tested is: "*there is no statistically significant relationship between the number of followers an account has, and the likelihood of that user remaining as part of the network*". The significance of results from a logistic regression using the number of followers as a variable are summarised in Table 6.12. This alternative hypothesis is based on the assumption that 'popular' accounts (i.e. ones with high numbers of followers) are likely to be relatively stable and/or selective in terms of which accounts they choose to follow.

Table 6.12: Hypothesis 6. Summary of statistical significance of results from logistic regression testing.

Location	$p<.001$	$p<.050$	$p<.100$	Mean p Value
Regent Street	1	3	4	.313
Carnaby London	0	1	1	.372
Seven Dials	1	3	4	.284
Oxford Street	0	2	2	.285
Covent Garden	1	4	5	.238
London West End	2	3	3	.313
Total	5	16	19	.301

In a similar vein to hypothesis 5, whilst all locations have at least one snapshot that indicates a statistical significance between the number of followers, and remaining a part of the network, it is a relatively low percentage across all locations. As such, here the null hypothesis is accepted - there is no consistent statistically significant relationship between the variables.

6.4.7 Hypothesis 7

Similar to hypothesis 1, this hypothesis is broken down into 4 sub-hypotheses. These hypotheses relate to the 'dynamic' calculated data for each user - that is, the change in the metrics between the previous and the current snapshot. Hypothesis 7A-7D relate specifically to the centrality metrics first tested in hypothesis 1. Where users are present in multiple snapshots, this data demonstrates whether the metrics are increasing or decreasing over time, and whether this change is eventually reflected in a user leaving the network, or remaining a part of the network. Each of these sub-hypotheses will now be tested in turn.

6.4.7.1 Hypothesis 7A

This hypothesis relates to the use of changes in the 'closeness centrality' metric as a means of predicting follower churn. It would be expected that an increasing centrality metric such as this would demonstrate a user becoming more highly connected and central to the network, and as such, they are likely to remain a part of the network. The null hypothesis - "*there is no statistically significant relationship between changes to a user's closeness centrality value, and their remaining part of the network*" will be tested, with the alternative hypothesis being "*users whose closeness centrality value is increasing are more likely to remain as part of the network*". To test this hypothesis, the difference in the closeness centrality between each snapshot and the previous snapshot was used as the independent variable in a logistic regression model. A summary of the statistical significance of the results for each snapshot, for each retail location can be seen in Table 6.13.

Table 6.13: Hypothesis 7A. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	
Regent Street	0	1	1	.424
Carnaby London	0	1	2	.390
Seven Dials	0	0	1	.544
Oxford Street	0	0	0	.580
Covent Garden	1	3	3	.346
London West End	0	1	2	.426
Total	1	6	9	.468

As can be seen in Table 6.13, very few of the locations' snapshots have significant results when using the change in closeness centrality to model and predict follower churn on social media. As such, the null hypothesis for 7A can be accepted - there is no consistent statistically significant relationship between the variables.

6.4.7.2 Hypothesis 7B

This hypothesis relates to the use of betweenness centrality - particularly the changes in the values of this metric, in order to model and predict follower churn: "*users whose betweenness centrality value is increasing are more likely to remain as part of the network*". As such, the null hypothesis being tested is "*there is no statistically significant relationship between a user's changes in betweenness centrality metrics, and their remaining within the network*". A summary of the statistical significance of these results can be seen in Table 6.14.

Table 6.14: Hypothesis 7B. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	Mean p Value
Regent Street	0	0	0	.779
Carnaby London	0	0	0	.665
Seven Dials	0	1	1	.568
Oxford Street	0	0	0	.710
Covent Garden	0	1	1	.513
London West End	0	1	1	.720
Total	0	3	3	.659

As can be seen in the results summarised in Table 6.14, there are very few occasions when the link between changes in betweenness centrality and remaining part of the network are statistically significant. In fact, only 3 of the 48 snapshots analysed have statistically significant results (p<.050) for this particular metric. As such, in this instance, the null hypothesis can be accepted, as there is no demonstrable significant relationship between the variables.

6.4.7.3 Hypothesis 7C

This (alternative) hypothesis relates to the third of the centrality measures selected for this study - transitivity: "*users whose transitivity measure is increasing are more likely to remain as part of the network*". Accordingly, the null hypothesis being tested is: "*there is no statistically significant relationship between changes in a user's transitivity metric, and their remaining part of the network*". To resolve this hypothesis, the change in transitivity values between each pair of snapshots was calculated, and then used as the independent variable in a logistic regression. A summary of the statistical significance of the results from these regressions can be seen in Table 6.15.

These results are inconclusive. While each of the retail locations has some results that are statistically significant, most retail locations have a low percentage of significant results. Only Covent Garden has a relatively high percentage (62.50%) of statistically significant results. Given this mix of results, the null hypothesis is accepted.

Table 6.15: Hypothesis 7C. Summary of statistical significance of results from logistic regression testing.

Location	$p<.001$	$p<.050$	$p<.100$	Mean p Value
Regent Street	0	1	1	.343
Carnaby London	0	1	3	.324
Seven Dials	0	1	3	.298
Oxford Street	0	1	2	.436
Covent Garden	1	5	7	.052
London West End	0	3	4	.342
Total	1	12	20	.299

6.4.7.4 Hypothesis 7D

In the last of these sub-hypotheses, changes in eigenvector centrality between snapshots are used as the independent variable in a binomial logistic regression. In this context, the null hypothesis being tested is: "*there is no statistically significant relationship between changes in a user's eigenvector centrality metric, and their remaining a part of the network*". A summary of the statistical significance of these results can be seen in Table 6.16.

Table 6.16: Hypothesis 7D. Summary of statistical significance of results from logistic regression testing.

Location	$p<.001$	$p<.050$	$p<.100$	Mean p Value
Regent Street	0	2	3	.307
Carnaby London	2	5	5	.140
Seven Dials	0	2	2	.363
Oxford Street	0	3	4	.247
Covent Garden	3	7	7	.031
London West End	8	8	8	$<.001$
Total	13	27	29	.217

From the significance of the results, summarised in Table 6.16, this is perhaps the most useful of these four metrics. This metric has statistical significance for all six of the retail locations, with three locations having significant results in over 50% of the analysed snapshots. As such, the null hypothesis is rejected, and the alternative hypothesis can be tentatively accepted - the results are significant in over 50% of the snapshots overall, and in almost all snapshots for both Covent Garden and London West End. With sufficient statistically significant results, it is now important to consider the size of the effect of this metric, and the goodness of fit of such a model. Tables presenting these results for each of the six locations can be found in Appendix C, with a summary of these presented here in Table 6.17.

Based on the results shown in Table 6.17, this model, using the changes in eigenvector centrality as the sole metric in predicting follower churn, is a better fit (indicated by a lower AIC value) than

Table 6.17: Hypothesis 7D. Logistic regression results table.

Location	Estimate	Std.Error	Null Dev.	Residual Dev.	AIC	R^2
Regent Street	10.3485	8.6806	1499.4	1496.7	1499.6	0.275800
Carnaby London	21.8458	6.9672	2399.5	2389.2	2393.2	0.210262
Seven Dials	10.8654	10.4901	1100.3	1098.7	1102.7	0.306209
Oxford Street	11.5864	8.6179	1666.5	1664.2	1668.2	0.261462
Covent Garden	16.2940	4.7883	3195.4	3186.1	3190.1	0.205850
London West End	27.6530	4.0378	3850.8	3802.5	3806.5	0.167964
Total	16.4322	7.2637	2285.3	2272.9	2276.7	0.237925

the metrics used for hypotheses 1C, 1D and 4. Further to this, the estimated effect size are, in general, higher than (for example) hypothesis 1D, with Nagelkerke's R^2 values all considerably higher. This would indicate that measuring the changes in the values over time is a more effective means of predicting follower churn than merely measuring the metrics at a single, fixed, point in time. This has implications for implementing such techniques - demonstrating that monitoring the development of such networks has demonstrable benefits.

6.4.7.5 Summary of Hypothesis 7

Hypothesis 7 was broken down into 4 distinct sub-hypotheses, each of which addressed the same core question, but utilised different centrality measures. Of these four, only a single metric - the changes in eigenvector centrality - was considered as a viable measure in order to predict social media follower churn. As discussed previously, this metric, however, was a much better means of predicting and modelling follower churn than the 'static' measures tested in previous hypotheses. As such, hypothesis 7 can be partially accepted - subject to the use of the correct centrality measure.

6.4.8 Hypothesis 8

This hypothesis makes use of the changes in measures between collected data snapshots. Here, the null hypothesis being tested is: "*there is no statistically significant relationship between changes in an accounts 'in-network' follower count and their remaining part of the network*". To test this hypothesis, the change in the number of 'in-network' followers is used as the independent variable in a logistic regression. A summary of the statistical significance of results of these tests is provided in Table 6.18.

The results presented in Table 6.18 clearly demonstrate that, in general, there is no statistically significant result when considering this hypothesis, as such the null hypothesis is accepted.

Table 6.18: Hypothesis 8. Summary of statistical significance of results from logistic regression testing.

Location	$p<.001$	$p<.050$	$p<.100$	Mean p Value
Regent Street	0	0	0	.562
Carnaby London	0	0	0	.667
Seven Dials	0	0	0	.669
Oxford Street	0	0	0	.877
Covent Garden	0	0	0	.731
London West End	0	1	1	.678
Mean	0	1	1	.697

6.4.9 Hypothesis 9

In this instance, the null hypothesis being tested is: "*there is no statistically significant relationship between changes in the number of 'in-network' accounts an account follows, and their remaining part of the network*". Here, the alternative hypothesis is: "*users are more likely to remain as part of the network if the number of 'in-network' accounts they follow is increasing*". This hypothesis can be tested using the change in the number of 'in-network' accounts that are being followed by each user. The statistical significance of the results from the logistic regressions are summarised in Table 6.19.

Table 6.19: Hypothesis 9. Summary of statistical significance of results from logistic regression testing.

Location	$p<.001$	$p<.050$	$p<.100$	Mean p Value
Regent Street	4	6	6	.107
Carnaby London	7	7	7	.031
Seven Dials	3	5	7	.117
Oxford Street	6	7	7	.064
Covent Garden	8	8	8	$<.001$
London West End	8	8	8	$<.001$
Total	36	41	43	.697

As can be seen from the results, there is a very high percentage of data snapshots that demonstrate a statistically significant ($p<.050$) relationship between this metric and whether users 'churn' or not. Accordingly, the null hypothesis is rejected, and the alternative hypothesis accepted. As statistical significance has been clearly demonstrated, it is now useful to consider the effect size, and goodness-of-fit of these models. This data for each retail location is shown in tables in Appendix C, with a summary of these shown in this chapter in Table 6.20.

The results in Table 6.20 demonstrate that the model using this particular metric is a better fit to the data, compared to hypothesis 7D, as indicated by lower AIC values. Further, the Nagelkerke's R^2 values demonstrate that using this metric accounts for more of the outcomes (i.e. remaining part of the network) than when modelled solely using the metric used in hypothesis 7D. Further, while

Table 6.20: Hypothesis 9. Logistic regression results table.

Location	Estimate	Std.Error	Null Dev.	Residual Dev.	AIC	R^2
Regent Street	0.0608	0.0147	1499.4	1485.9	1489.9	0.282854
Carnaby London	0.0585	0.0097	2399.5	2346.1	2350.1	0.227561
Seven Dials	0.0683	0.0206	1100.3	1084.2	1088.2	0.317936
Oxford Street	0.0902	0.0195	1666.5	1638.3	1642.3	0.272377
Covent Garden	0.1175	0.0115	3195.4	3041.0	3045.0	0.250118
London West End	0.1733	0.0098	3850.8	3308.1	3312.1	0.283869
Mean	0.0948	0.0143	2302.0	2150.6	2154.6	0.272619

the standard error values for this metric are also very low, this is due to both the small values being handled when using this metric, and also the small effect sizes being encountered. These results demonstrate that, as users begin to follower more accounts that also follow the main account, they are less likely to churn. This would suggest that individuals being increasingly connected in the network results in them remaining a part of the network. Conversely, if users were seen to be becoming less connected within the network, then they would be at risk of churning in the future.

6.4.10 Hypothesis 10

Hypothesis 10, as previously outlined, is "*users are more likely to remain as part of the network if their number of followers is increasing*". The null hypothesis, therefore, is that "*there is no statistically significant relationship between changes in the number of followers an account has, and their remaining part of the network*". This hypothesis can be tested using the number of followers of each account as the independent variable in logistic regression models. The statistical significance of the results from these can be seen in Table 6.21. From these results, it can be seen that, in general, p values tend to be high, with only four of the six retail locations demonstrating any significant ($p<.050$) results, and these are only seen in a low percentage of the analysed snapshots. For this reason, the null hypothesis is accepted - there is no demonstrable statistically significant relationship between the variables.

6.4.11 Hypothesis 11

This hypothesis tests the use of multiple metrics to model the likelihood of users remaining part of the network. From the previous ten hypotheses, the following metrics are shown to have statistically significant uses when modelling follower churn: transitivity, eigenvector centrality, whether a user is also following others in the network, an increasing eigenvector centrality, and an increase in the the 'outdegree' of a user - that is, the number of 'in-network' accounts that they also follow.

Table 6.21: Hypothesis 10. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	Mean p Value
Regent Street	0	1	1	.640
Carnaby London	0	0	0	.548
Seven Dials	0	0	0	.628
Oxford Street	0	1	1	.524
Covent Garden	1	2	2	.524
London West End	1	3	3	.458
Total	2	7	8	.564

Both transitivity and eigenvector are both centrality measures, and changes in the eigenvector centrality measure have been shown to be a more effective model than the 'static' measures of either eigenvector centrality or transitivity. Further to this, the changes in the number of 'in-network' accounts being followed has been shown to be far more effective a model than merely whether or not the user also follows 'in-network' accounts.

Therefore, for this particular hypothesis, a combination of the two metrics: changes in eigenvector centrality, and changes in the number of 'in-network' accounts being followed (outdegree) will be used. As such, the null hypothesis being tested is: "*there is no improvement in the models when using multiple variables to predict if a user will 'churn'*". The results from these models can be seen in Table 6.22, with results for each location detailed in Appendix C.

From the earlier hypotheses, hypothesis 9 produced the more favourable results, in terms of goodness of fit of the model to the data. Comparing the results from this model indicates that including multiple metrics (in this case changes in eigenvector centrality, and changes in the number of 'in-network' accounts being followed) produces marginally more favourable results. This can be seen through lower residual deviance values, AIC values that are either lower than, or comparable to those produced in hypothesis 9, and Nagelkerke's R^2 values that are slightly higher.

Although the change in the results is not particularly large, here the null hypothesis can be rejected and the alternative hypothesis accepted - using multiple metrics to model and predict social media follower churn does produce more effective results than using a single metric.

Table 6.22: Hypothesis 11. Logistic regression results table.

Location	Null Dev.	Residual Dev.	AIC	R^2
Regent Street	14994.	1359.0	1490.0	0.283762
Carnaby London	2399.5	2267.3	2352.0	0.227660
Seven Dials	1100.3	1083.7	1089.7	0.318300
Oxford Street	1666.5	1636.4	1631.1	0.272736
Covent Garden	3195.4	3039.3	3045.3	0.250637
London West End	3850.8	3295.1	3301.1	0.286201
Mean	2285.3	2233.5	2151.5	0.273216

6.5 Discussion of Results and Implications for Future Work

In this section, the implications of these results are discussed in greater detail. First, the use of social network analysis metrics to predict and model follower churn is summarised. Following this, how these findings may be used to inform future work is also discussed.

6.5.1 Using Social Network Analysis Metrics to Model Follower Churn

In hypothesis 1-10 (and their sub-hypotheses), the use of individual social network analysis metrics in predicting and modelling follower churn was tested. While in many cases, the null hypotheses were accepted due to the varied and statistically insignificant results, some of the null hypotheses were rejected, and the alternative hypotheses accepted.

Hypotheses 1C and 1D both related to centrality measures at any given point in time. Both of these indicated some usefulness in modelling and predicting follower churn. However, the results also indicated that such metrics could only be used to account for a fairly small percentage of the outcome - suggesting that other factors were still effecting the likelihood of users churning.

Hypothesis 4, which relates to who the user is also following, also produced similar results - while using this metric is useful to some extent, the results demonstrated that they do not account entirely for the outcome (i.e. if the user remains in, or leaves, the network).

Hypothesis 7D and 9 (in both cases the null hypothesis was rejected, and the alternative hypotheses accepted) make use of 'dynamic' metrics - that is, changes in the values from one snapshot to the next. While the previous hypotheses made use of statistics indicating the users' positions at a given point, these hypotheses made use of the changes in these values from one hypothesis to the next, thus indicating the development of each individual's position within the network. Both of the metrics used to test these hypotheses are demonstrably better than using the 'static' metrics used in previous hypotheses, with Nagelkerke's R^2 metrics that are far higher than the results from the testing of previous hypotheses.

From these, it can be deduced that while making use of single social network analysis metrics produces useful results, and can help in modelling and predicting social media follower churn, tracking these metrics over time and using changes in these values produces more useful results. These findings begin to address the need for understanding the predictability of online behaviours, identified from existing literature. Future work could confirm these findings with more longitudinal studies and including more accounts. However, these findings indicate that social network metrics can be used (to some extent) to predict online behaviour and that this use is strengthened when metrics are tracked and recorded over time.

Hypothesis 11 relates to the use of a combination of the previous metrics to model and predict social media follower churn. The results from this particular test demonstrate that using a combination of metrics produces slightly more favourable results than any single metric that had been previously tested. This suggests that, rather than focusing on a single metric that is the ideal predictor of follower churn, developing groups of metrics and indicators will be more useful in modelling and predicting social media follower churn. The Nagelkerke's R^2 values also suggest, perhaps, that models based solely on an individual's position within a network will be limited in their effectiveness, and may require other factors to be taken into consideration.

6.5.2 Informing Further Work

The findings presented in this chapter have clearly demonstrated that, from the data collected and analysed, social network analysis techniques and associated metrics can be used to model and predict social media follower churn. As such, this approach has merits for organisations wishing to begin to identify those accounts that may be at risk of churning.

In doing so, this study has contributed to addressing some of the previously discussed motivations for research, such as the need for longitudinal studies in this field [60], understanding sharing behaviours [171], and the motivations for users to engage with content [60]. By implementing a longitudinal study, as suggested by prior work [60], this study was able to consider network churn over a much longer period of time, and include multiple 'snapshots' of the status of each network. This allowed for more data points to be used in the statistical modelling, and as such strengthens the validity of the observed results, in comparison to other works, or if this study had been conducted over a much shorter period of time, with fewer 'snapshots' of data. While providing useful and beneficial findings, this study has also demonstrated that further work is still required in this area.

What these results have also demonstrated is that these metrics should not be used on their own to reliably model and predict social media follower churn. Some of the tested metrics - namely those that measure the changes in the values over time - have been shown to be more effective at modelling the outcome (i.e. if a user remains in or leaves the network). However, all of the metrics with statistically significant results have demonstrated that there is far more that could be done to accurately model this phenomenon. The metrics that produced the highest Nagelkerke's R^2 values still only resulted in figures of around 0.2 to 0.3, suggesting that there are other factors at work, that (if possible) could be analysed and included in these models.

6.6 Summary

In this chapter, a study has been presented which tests and demonstrates the use of social graph metrics to model and predict social media users that are at risk of churning - that is, at risk of unfollowing the social media account being focused on.

Through a series of logistic regressions, each of these hypotheses have been tested. First, individual metrics were tested, and those that were deemed to have sufficiently statistically significant links were explored in more detail. Following this, the final hypothesis related to the use of multiple metrics to model the likelihood of individual social media users churning.

From the results gained from these tests, it has been demonstrated that there are multiple metrics that have statistically significant links between a social media user's position within a social network, and whether or not they are likely to churn. Narrowing these results down and including two metrics - changes in eigenvector centrality, and the 'outdegree' of a user (i.e. the number of 'in-network' accounts that they follow) result in a model with a marginally higher Negelkerke R^2 value, suggesting a model with a somewhat better ‘fit’ to the data.

These results demonstrate that while there is some value in using these metrics in isolation to predict and model follower churn, that more could be done to supplement this approach, and other factors taken into account as part of this process. One such feature that could supplement this approach is engagement with existing social media content. This will be explored in more detail in the following chapters. In the next chapter, a technique for developing engagement profiles for social media users is demonstrated and evaluated.

Chapter 7

Modelling User Engagement with Promotional Social Media Content

7.1 Introduction

Previous chapters have demonstrated the use of social media by retail locations (Chapter 5), and the way in which the networks around these social media accounts can grow or decline (Chapter 6). What became apparent from the work presented in both Chapter 5 and Chapter 6 was that retail locations seem to employ different strategies in terms of types of content they share on social media, but also that individuals respond to these types of content differently. From the results presented in Chapter 5 we also see that this results in differing levels of engagement with this content, that may contribute to the growth or decline of follower networks, as presented in Chapter 6.

The study presented in this chapter demonstrates and evaluates a new method of generating, and grouping, user profiles based on their engagement with types of content posted by specific social media accounts. Such groupings could be looked upon as another means of conducting market segmentation, with a specific focus on engagement behaviours. The ultimate goal of such an approach is to facilitate a greater understanding of an account's followers and the content with which they are likely to engage. This would then allow for the creation of useful, targeted, social media content that would encourage higher levels of engagement. Not only would such engagement propagate the organisation's posts further, but previous research suggests that engagement with such content has more long-term benefits for customer retention, loyalty and purchasing behaviours [6, 75].

A number of research contributions are demonstrated within this chapter:

- The method for creating social media user profiles, based on qualitative analyses of the content with which they have publicly engaged - a method which can reduce large, and often complex, datasets to a series of vectors representing engagement behaviours.
- Demonstrating that communities of users identified using widely-used social network analysis techniques do not engage with social media content in the same way.
- The results demonstrate that the way in which the majority of social media users engage with

social media content from organisations and brands does not vary dramatically over time.

- Suggestions for how this information can be used in the creation of effective social media strategies for organisational Twitter, and other social media, accounts are provided – including the practicalities of adapting and integrating this method within the content creation process.

This chapter is structured as follows. First, a description of the selected retail locations, which act as case studies, is provided. Following this, the process by which the collected data was analysed, and engagement profiles constructed, is discussed. The proposed method of clustering engagement profiles is then explored in detail, including a comparison of these resultant clusters to communities of users within the network, that were identified using social network analysis techniques. An analysis of the validity of these engagement profiles over time is then followed by a brief comparison of the proposed analysis techniques to those offered by 'off-the-shelf' solutions. The chapter then concludes with an evaluation of the approach, as well as a discussion of the implications for how this process may be used and developed further.

7.2 Selection of Retail Locations as Case Studies and Data Overview

To address the goals of the study presented here, two Twitter accounts were selected to act as case studies. Twitter offers useful public insight into individuals' engagement behaviours, through the use of the 'retweet' function, which propagates the original message to that individual's followers, and as such is generally considered to be a positive outcome for the account responsible for posting the original message.

The two Twitter accounts selected (those of Regent Street, London; and Intu Metrocentre, Gateshead), motivated by ongoing work with industry partners, were selected as they both represent a retail location which comprises many individual tenant retailers. As such, the accounts are responsible for not only promoting the location itself, but also the interests of the individual tenant retailers. These locations are geographically distinct, with Regent Street being a high-value shopping district in the west end of London, and intu Metrocentre being a large out-of-town shopping mall, near Newcastle-upon-Tyne in the north-east of England. Not only do these accounts represent high-profile, high-footfall retail locations, but they are actively maintained by their respective public relations and media teams, and as such present ideal case studies for this study.

Three types of data were required for this study: public tweets sent from the accounts being studied, a list of all public users that engaged with these tweets, and the social graph data for the networks surrounding both of the accounts being studied. Each of these is described below, and the datasets themselves are summarised in Table 7.1.

First, it was necessary to collect the public tweets sent from each of these accounts, over period of time covering a total of 18 months. The first twelve months of this form the 'original' dataset. The 'validation' dataset consists of data covering a period of three months, which itself begins three months after the end of the original dataset. While the use of these two datasets is explained in greater detail later in this chapter, the break between the original and validation datasets was to allow for time to pass, and the audience to develop before validating the techniques and approaches used in the study.

Table 7.1: A summary of the datasets collected and used to construct and evaluate social media user engagement profiles.

Dataset	Relative Dates	Size of Dataset
All public tweets sent from the accounts in a 12-month period, referred to as the 'original datasets'	Month 1 - Month 12	5,815 tweets
All public tweets sent from the accounts over a 3-month period, beginning 3 months after the end of the original datasets, referred to as the 'validation datasets'	Month 16 - Month 18	969 tweets
A list of all users that have retweeted any of the tweets in both the original and validation datasets	Month 1 - Month 18	1,062 users
The complete social graph of all the public followers of the Twitter accounts	Month 1 - Month 18	17,943 users

Public tweets sent from each of the two accounts were collected using Twitter's User Timeline API endpoint. At the time of data collection, Twitter imposed a limit to the number of tweets that could be collected using this endpoint - approximately 3,200 tweets. However, the total number of tweets sent from each account during the 18-month time period being studied was lower than this number. As such, no issues were encountered due to this API limitation.

It was also necessary to collect a list of each user ID that had engaged (i.e. retweeted) with each of the collected posts. By iterating through each collected tweet, and making use of the 'Statuses / Retweeters' API endpoint, it was possible to collect this data. Users whose accounts had been deleted, or suspended, since retweeting a post would not be included in this data, nor would users whose accounts are marked as 'private'.

Finally, it was necessary to collect data in order to construct the social graph of the followers of each of the two retail location's Twitter accounts. In order to collect this data, the 'Followers' and 'Friends' API endpoints were used. The Followers endpoint returns a list consisting of all of the public accounts that follow a specified account. The Friends endpoint returns a list of all of the public accounts that a specified account follows. First, the Followers endpoint was used to collect the IDs of each account that followed the two locations being studied. In order to construct the links between these follower accounts, the Friends API was used. Although the equivalent data could be collected using the Followers endpoint, the Friends endpoint was selected as the dataset included some high profile retailers who have significantly higher follower counts than they do friends counts (i.e. they could be followed by millions of accounts, but only follow 100 accounts). As these Twitter API endpoints are rate-limited, by collecting the data in this way, the number of required API calls was reduced considerably, and as such, the data collection time was drastically reduced.

As the details of private, suspended, or deleted accounts are not made available through these public API endpoints, any such accounts did not form any part of these datasets. As such, the total number of accounts included in the analyses presented in this chapter is likely to be lower than the published number of followers of each of these accounts at the point of data collection. While this may exclude them from use in this study, and prevents (for example) the characterisation of their engagement behaviours, private accounts are also limited in the extent to which they could help to propagate or support an organisation's message. As such, other than reducing the amount of available data for this study, it does not have a dramatic impact on the results of this study.

7.3 Qualitative Analysis of Tweet Content

To characterise the content of each of the collected tweets - in both the original and validation datasets - and thus enable the creation of users' engagement profiles, each tweet was subjected to a manual, qualitative, analysis. As detailed in previous chapters, the methods utilised by Braun and Clarke [27, 28, 29, 30] were used.

As part of this approach, codes (or labels) were applied to each of the collected tweets, indicating the types of content in each tweet. The codes were then re-checked, and refined where necessary. This process led to a final list of 20 content codes, which are listed in Table 7.2. The number of codes applied to each tweet differed on a case-by-case basis, with some tweets having multiple codes applied, although most had only a single code applied to them.

At this point, it was possible to begin to group codes together in order to form different themes. However, for the purposes of creating engagement profiles (detailed previously in Chapter 4, and also covered later in this chapter), it was decided to use the content codes, rather than themes. As such, no themes were formed and the qualitative process was concluded at this point.

Table 7.2: List of codes applied to tweets during the construction of user engagement profiles

#FollowFriday	Offer Information
Asking a Question	Personalised
Asking for a RT	Product Comment
Calendar Events	Promoting an App
Commenting on Shared Media	Promoting Social Media
Competition Information	Prompting for Response
Directed or Conversational	Retailer Comment
Event Comments	Sharing an Image
General Promotion of Location	Sharing a Link
Manual Retweet	Weather Information

7.4 Construction of Engagement Profiles

With the content of each tweet labelled, it is then possible to associate this information back to the individual users that engaged with each of these tweets, in order to create each user's engagement profile. In this instance, only data from the original dataset was used.

The engagement profile is represented by an n-dimensional vector, with each dimension representing one content code from the analysed dataset. For example, in this particular case study, each engagement profile vector would consist of twenty dimensions, with one representing 'Asking a Question', another 'Calendar Events' etc. For each dimension of the vector, the value ranges between 0.00 and 1.00, representing the percentage of each user's engaged-with tweets that contained that particular content code. For example, a user who engaged with a total of three tweets, two of which were categorized as 'Sharing an Image', would have a value of 0.66 in the 'Sharing an Image' dimension. This process is demonstrated in Figure 7.1.

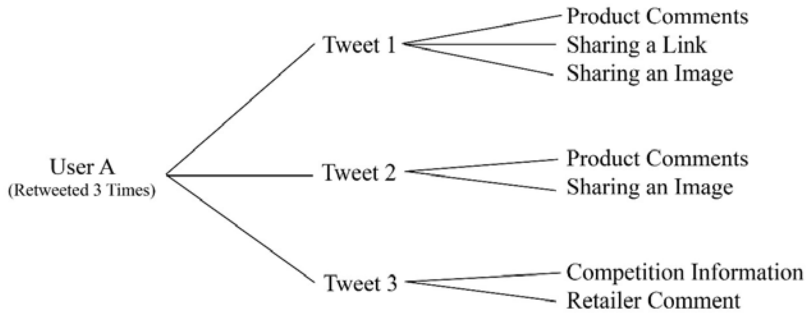


Figure 7.1: A visual demonstration of creating engagement profiles, linking content codes from analysed tweets to followers that have engaged.

It should be noted, at this point, that these vectors are based solely on the number of tweets with which a user has been seen to publicly engage. Due to the substantial number of posts that may be presented to a user in their timeline at any given time, it cannot be assumed that users have seen each of the analysed posts and made a conscious decision to not engage with it. However, modelling their engagement in this way (i.e. as a share of the number of posts that they can be seen to have engaged with) reduces this uncertainty. Each profile can, therefore, be read as 'of the total number of tweets they have engaged with, they engage predominantly with this content type'. Modelling their engagement as a percentage of the total number of tweets sent from the analysed accounts would be problematic, as there is no way of guaranteeing that they have seen each of those tweets.

Such an approach to modelling user engagement can be adapted easily to suit the context in which it is being used. For example, while this analysis resulted in the generation of twenty codes, other analyses on other datasets is likely to result in a different number. Further to this, the analysis itself can be adapted to other platforms, engagement metrics, and content types.

7.5 Clustering Engagement Profiles

Aggregating this collection of engagement profiles into clusters of 'similar' users - that is, users that engage with similar types of content, will be beneficial and offers an alternative means of market segmentation. These clusters will demonstrate not only the types of content that are often engaged with, but also the size of the clusters will indicate the share of the audience engaging with the different types of content.

As a topic, data clustering spans multiple disciplines and areas, with many different approaches to clustering being proposed in literature [100]. In this particular study, k-means clustering was selected as a suitable approach - this is a widely-used algorithm for the low-cost clustering of multivariate data [100]. As part of this clustering algorithm, the value of k (i.e. the number of clusters) needs to be pre-determined. Again, there are multiple means of doing so, including the gap statistic, which was utilised in this instance [189]. The final choice required for this approach is the distance metric to be used. Euclidean distance is widely used in clustering methods, and produces favourable results when used with datasets with a high number of dimensions. As such, it was deemed to be an appropriate choice for this study, not only as it is suitable for the case studies presented here, but will also be suitable if used in other contexts with higher numbers of content codes (and thus, a higher number of dimensions), for example.

Once the clustering algorithm had been applied to the data, it was then possible to characterise the *general engagement behaviours* of each cluster. This was done through the use of the centre of each cluster to indicate the types of content typically being engaged with by the users assigned to each cluster. The results for both of Twitter accounts being studies were based around a combination of 8 different themes of engagement, which have been derived from the content codes at the centre of each cluster.

- *Competition Engagers*. Here, users are often engaging with content that was related to competitions being run, or advertised, through the Twitter account.
- *Event Engagers*. In these clusters, users were likely to engage more with content that was related specifically to events being advertised and discussed.
- *Information-Focused*. Users in these clusters often engaged with content that was deemed to be sharing further information, not just contained within the tweet, such as links to external websites.
- *Product-Focused*. Users were more likely to engage with content that referenced either a single product, or a group of products.
- *Relational*. Users in these clusters were more likely to engage with posts that were either personalised (perhaps mentioning the user directly), or conversational in tone.
- *Retailer-Focused*. In these clusters, users were engaging with posts that promoted a specific retailer, or group of retailers.
- *Self-Promoting*. These clusters contained users that were more likely to engage with tweets that were related to themselves. Such tweets include those that commented on media the user had shared online, or perhaps relating to competitions that they had entered or won.
- *Visual*. These users more readily engaged with visual content - such as images or video clips.

Some of these labels were assigned to multiple clusters, but each cluster represents a unique combination of the types of content being regularly engaged with.

The results of the clustering algorithm generated a solution of eleven user clusters for Regent Street (with cluster size varying between 7 and 121 users), and ten clusters for Intu Metrocentre (with clusters varying in size between 5 and 97 users). As outlined previously, using the centre of each of these clusters allows for a general profile to be generated for each cluster, indicating the types of content most often engaged with by users in that cluster. The main engagement behaviours of each cluster, and the size of each cluster, can be seen in Table 7.3 for Regent Street, and Table 7.4 for Intu Metrocentre.

Table 7.3: Main engagement behaviours for each engagement cluster for the Regent Street Twitter account, also showing the size of each engagement cluster.

#	Main Behaviours	Size
1	Information-focused	56
2	Competition engager	82
3	Visual, retailer-focused	85
4	Information-focused, event engager	59
5	Event engager	121
6	Visual, event engager	43
7	Retailer-focused, event engager	28
8	Event engager, relational	65
9	Competition engager, self-promoting	7
10	Information-focused, competition engager	38
11	Visual, product-focused	32

7.6 Comparison of Engagement Clusters and Social Network Analysis Communities

One of the broader aims of this study was to compare the results between two different approaches to understanding social media users - the clusters of users that engage with similar types of social media content, and communities of users identified using social network analysis techniques. If the two methods produce similar results, then it would suggest that strongly-tied communities of users may engage with similar types of social media content. If not, then it suggests that solely relying on SNA techniques for grouping users together is unlikely to produce effective results in this context.

From the collected social graph data (summarised in Table 7.5), communities of users were determined using the techniques detailed in [19, 143, 144, 145]. This generated results indicating 7 communities of users in the Regent Street social graph, and 10 communities of users in the Intu Metrocentre social graph. The results and comparisons for both Regent Street and Intu Metrocentre will now be explored and discussed in turn.

Table 7.4: Main engagement behaviours for each engagement cluster for the Intu Metrocentre Twitter account, also showing the size of each engagement cluster.

#	Main Behaviours	Size
1	Relational	25
2	Competition engager	73
3	Information-focused, event engager	97
4	Competition engager, information-focused	19
5	Product-focused, visual	5
6	Event engager	22
7	Information-focused	18
8	Event engager, visual	35
9	Retailer-focused	62
10	Event engager, competition engager	87

Table 7.5: An overview of the properties of the constructed social graphs for the Regent Street, and Intu Metrocentre Twitter accounts.

	Regent Street	Intu Metrocentre
Nodes	6,388	11,555
Edges	121,408	525,526
Number of Communities	7	10
Modularity	0.36	0.27
Network Diameter	8	7
Graph Density	0.003	0.004

7.6.1 Regent Street

A summary of the results for Regent Street can be seen in Table 7.6. This table demonstrates the number of accounts from each cluster that belong to each of the social graph communities, and vice versa. The majority of engagement clusters are dispersed across most, if not all, of the the social graph communities. The one major exception to this is cluster 2, which shows over 95% of its members assigned to social graph community 7. Despite this, in general terms, there is no clear-cut relationship between engagement behaviour and social graph community. This would suggest that identified communities with a social network do not necessarily engage with content from a specific account in the same ways.

7.6.2 Intu Metrocentre

In a similar vein to the results for Regent Street there is, for the most part, no clear and defined relationship between the engagement behaviours of individual social media users, and the social graph community to which they are deemed to belong.

Table 7.7 summarises the results for this Twitter account. While the presented results demonstrate a clear majority of users of cluster 5 belonging to social graph community 10, this result is likely to

Table 7.6: Breakdown of RegentStreetW1's SNA communities and engagement cluster membership. Table shows total cluster sizes, as well as the number of accounts from each community that have engaged.

	SNA Community							
Cluster	1	2	3	4	5	6	7	Total Size
1	12	10	6	15	6	3	4	56
2	0	1	1	0	1	1	78	82
3	8	18	29	12	6	2	10	85
4	6	8	7	21	5	2	10	59
5	18	27	15	14	16	17	14	121
6	3	13	10	7	8	1	1	43
7	2	7	7	5	5	1	1	28
8	4	10	14	18	14	4	1	65
9	0	2	3	1	1	0	0	7
10	2	9	7	3	2	4	8	35
11	4	6	4	4	2	4	8	32
Total Engaged	59	111	103	100	66	39	135	

Table 7.7: Breakdown of intoMetroCentre's SNA communities and engagement cluster membership. Table shows total cluster sizes, as well as the number of accounts from each community that have engaged.

	SNA Community									
Cluster	1	2	3	4	5	6	7	8	9	10	Total Size
1	0	4	4	0	1	6	0	0	2	8	25
2	1	1	28	0	6	6	0	0	3	28	73
3	5	18	12	3	23	12	0	0	4	20	97
4	1	1	5	0	5	0	0	0	0	7	19
5	0	0	0	0	0	1	0	0	0	4	5
6	0	2	4	0	2	0	0	0	5	9	22
7	0	2	0	0	2	3	0	0	2	9	18
8	2	3	4	0	7	7	0	1	1	13	38
9	1	8	14	1	3	10	0	0	4	21	62
10	3	1	37	0	6	5	0	0	1	34	87
Total Engaged	13	40	108	4	55	50	0	1	22	153	

be skewed by the fact that there are only five followers within that engagement cluster. Again, cluster 7 demonstrates a strong relationship between the two, with half of its eighteen member users being assigned to the same social graph community.

The remaining engagement clusters, especially those with larger amounts of users assigned to them (such as clusters 2,3,9, and 10) show their users dispersed across various social graph communities. It is interesting to note, however, that some communities do not engage at all - such as community 7 - or have very few members that engage (i.e. communities 4 and 8). Information such as this may well be of use to social media account managers when implementing and developing their social media strategies. Such information would not necessarily have been evident if the engagement behaviours and preferences of these communities had not been considered in this way.

7.6.3 Summary of Comparisons

The results for both accounts that have been used as case studies demonstrate that there is no clear, well-defined relationship between the social graph community of a user (indicating the group of users to which they are more heavily connected) and their general engagement behaviour in terms of the types of content with which they are seen to engage. Some of the results suggest that there may be a correlation between the two, but these are for particularly small groups of users, which may skew the results.

These results, particularly for Intu Metrocentre, have demonstrated that there may be some communities of users within the accounts' follower network that do not engage at all. While this study does not explore the potential reasons for this, such information is likely to be of use for social media account managers, as they would be able to identify whole groups of users that have not engaged previously. How they act on this information is likely to change from sector to sector, and from organisation to organisation. While some may choose to disregard these users, others may look at ways of encouraging engagement from that group of their followers.

7.7 Validity of Engagement Profiles Over Time

Whilst previous sections of this chapter have proposed and demonstrated a technique for developing engagement profiles for social media users, any such profiles would need to be valid - and therefore useful - for a period of time after they are created. By ensuring that these results are still valid, social media campaigns and strategies can be reliably created, based on these results.

In order to validate the longevity of these engagement profiles - the content codes applied to the validation dataset (see a summary of datasets in Table 7.1) were used to generate engagement profiles for each of the users that have engaged in this second time period. These profiles were then compared to those generated previously, based on the original dataset. Two different measures were selected to aid in the validation of these profiles: the Euclidean distance between the two engagement profiles - original and validation; and changes in the general engagement behaviours of each individual user. Both of these approaches are detailed in turn, in the following sub-sections, with the results for both locations, Regent Street and Intu Metrocentre, discussed.

7.7.1 Profile Similarity using Euclidean Distance

Euclidean distance can be used to measure the distance between two vectors in n-dimensional space, using the formula demonstrated in Equation 7.1.

$$Euclidean\,Distance = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2+...+(p_i-q_i)^2+...+(p_n-q_n)^2} \tag{7.1}$$

To measure the variation between the original and validation profiles, the Euclidean distance between the two was calculated for each user that exists in both the original and the validation datasets. Measuring the change in this manner will provide equal weighting to variations in each of the content-type categories. In other contexts, it may be deemed that some content types are of higher importance than others, and the technique could be adjusted accordingly. To help contextualise the difference between the two sets of profiles, the results are represented as a percentage of the maximum possible variation between the two profiles. For example, a 100% variation would demonstrate a complete reversal across all types of content - where a user had engaged with 100% of tweets of a particular content type, they would engage with 0% of the same type in the validation profile, and vice versa.

7.7.1.1 Regent Street

Of those users that engaged with content from both the original and validation datasets, the minimum demonstrated variation was just 5.56% of the maximum possible distance, with a maximum variation of 48.99% ($\bar{x} = 26.86\%$, $\tilde{x} = 26.00\%$). If one-time engaging users are omitted for this data, this range in variation becomes 5.56% to 31.20% ($\bar{x} = 20.35\%$, $\tilde{x} = 20.96\%$).

7.7.1.2 Intu Metrocentre

For Intu Metrocentre followers, the minimum variation for engagement profiles between the original and validation datasets was 13.58%, with a maximum variation of 50.00% ($\bar{x} = 33.97\%$, $\tilde{x} = 36.40\%$). In comparison to the Regent Street dataset, a higher percentage of these users (75%) only engaged once in the time period covered by the validation dataset. If these users are omitted from the results, the variation in profiles becomes 13.58% up to 40.54% of the maximum possible variation ($\bar{x} = 29.60\%$, $\tilde{x} = 31.27\%$).

7.7.1.3 Relationship between Engagement Levels and Profile Accuracy

When using this metric to measure the changes in individual users' engagement profiles, it is clear than no single user remains exactly the same in the types of content with which they do, or do not, engage. It would be highly unlikely that any user would have maintained an identical set of engagement behaviours over this period of time. When one-time engagers are removed from the analysis, the mean and median values both drop - indicating that (as one would expect) engagement profiles based on more data are more accurate over time. A strong, negative correlation exists between the amount of times a user engages with the Twitter account, and the level of variation in their engagement profile over time ($r_s = -0.606$).

7.7.2 Profile Similarity Using General Engagement Behaviours

As detailed previously, the centre of the engagement clusters has been used to indicate the *general engagement behaviours* of the users within that cluster. This highlights the key content types with which they have been engaging. By comparing these behaviours in the original and validation datasets, it can be shown how, and to what extent, the engagement behaviour of a user may change over time. Again, the results for both Regent Street and Intu Metrocentre will be discussed in turn.

7.7.2.1 Regent Street

For those users that engaged with content in both the original and validation datasets, a total of 46.58% of users demonstrated identical general engagement behaviours within both datasets. That is to say, that the dominant content types with which they engaged remained exactly the same - no general engagement behaviours were added to, or removed from, their profile. Further to this, another 32.88% of users retained the same engagement behaviours, but also demonstrated an additional behaviour. As such, for a total of 79.46% of these users, their original engagement behaviours were still evident in the validation dataset.

The remaining 20.54% of users in this dataset lost at least one of their prevalent engagement traits between the original and validation datasets. However, each of these users only engaged once during the validation dataset time period. As has been discussed previously, there is a correlation between the amount of data used to construct these engagement profiles, and the variation in observed behaviour between the two datasets.

7.7.2.2 Intu Metrocentre

These results are similar to those for Regent Street. 43.75% of users present in both the original and validation datasets for Intu Metrocentre demonstrated identical general engagement behaviours. A further 34.38% of users presented the same behaviours, with the addition of a further trait in the validation dataset. This results in a total of 78.13% of users who appear to sustain their engagement behaviours between both datasets, similar to the Regent Street value of 79.46% of users.

7.7.3 Summary

Two distinct methods of indicating the validity of these engagement profiles over time have been proposed, and evaluated. When using Euclidean distance to measure the variation in profiles for users, average variation in the profiles tended to be around 30%. However, this method treats each of the content codes equally, and as such there may be influence from some content codes that the users, the account managers, or both, may deem peripheral to the main content types being engaged with. When considering the broader, general engagement behaviours of users, a majority of users maintain the same, or very similar, engagement behaviours between the original and validation datasets.

These results would suggest that profiling users engagement using this technique produces results that are valid over a prolonged period of time. As such, social media campaigns and strategies could be developed around the engagement profiles of known groups of users - as opposed to creating content based on intuition and best-guesses.

7.8 Comparison of Method to Automated Analysis Tools

Previous sections have demonstrated a method for creating, and grouping, engagement profiles for individual social media users; further, the validity of these profiles over time has also been demonstrated. Part of this process involved the qualitative analysis of tweet content, in order to generate the user engagement profiles. While, in this case, this has been carried out ad-hoc (and thus may seem to be excessively time consuming), this process could be integrated into the business processes of the organisations making use of this process - this is discussed in more detail later in this chapter.

In order to streamline this process, some organisations may be tempted to make use of automated analysis tools, which are often readily available online (often at little or no cost for the 'basic' version), and which highlight key words, phrases, or concepts that are present in the text being analysed. In doing so, organisations may seek to save time, and hope to achieve similar results to the manual processes used within this chapter.

To provide a further basis for evaluation and comparison, and to further justify the resources required for such an approach, the same datasets have been processed using readily available text analysis tools. An overview of the results provided by these tools is provided, and then compared to the results gained through a manual analysis.

7.8.1 Availability of Automated Tools

As discussed previously, a range of online text analysis tools is available to marketers and researchers alike, with many offering a similar selection of features: keyword and concept identification, sentiment analysis, language detection, and in some cases, the 'ability' to identify and classify personalities and personas. These tools can be appealing to organisations as they are seen to offer a low cost - both temporal, and financial - entry to understanding, often on a very basic level, the contents of large-scale, text-based datasets. The question then arises, however, as to the quality and relevance of results generated by tools such as these, particularly in the context of the process detailed in this study. Further increasing their usability in both corporate and research environments, many of these have thorough documentation, competitive pricing strategies (including free entry-level features), codebases that can be integrated easily into existing products, and dedicated technical support.

7.8.2 Results from Automated Analysis Tools

In order to compare the results generated through the manual analysis proposed here, and those generated by readily available online tools, two of these tools were selected to act as a case study. Both Alchemy API [5], and Text Razor [188] were selected due to both having each of the functions listed in the previous section of this chapter, as well as advertising that they were suitable for use with small samples of text - such as tweets. While these are just two examples of a wide range of similar services and platforms, they serve as a demonstration of the type of results that can be gained through the use of such platforms with text-based Twitter content.

The same dataset using earlier in this study was processed using both of these tools; first, the 'keyword' methods were used, and then the 'concept identification' methods [5, 188]. Concept identification provides a further level of abstraction from keywords, grouping related terms together, and providing a higher-level view of the analysed content. Keywords identified using these tools were too numerous and fine-grained to be useful in the context of this study and the proposed approach. The most relevant and common concepts identified from these tweets, for both Regent Street and Intu Metrocentre's Twitter accounts, are listed in Table 7.8.

Table 7.8: A summary of results gained from two text-analysis tools, for both Regent Street and Intu Metrocentre data.

Account	Text Razor	Alchemy API
@RegentStreetW1	Regent Street; Facebook; Gift Card; Social Media; London; Apple Store; Twitter; Sweepstakes; Prize; Breaking News; Apple; Flagship; YouTube; Blog; Soundcloud	English Language Films; Giving; Earth; World; 2006 Albums; Google; Sociology; Luck; Apple Inc; Apple Store; Steve Jobs; Superstition; Internet Culture; Parties; Windows
@intuMetroCentre	Christmas; MetroCentre; TopShop; House of Fraser; Elephant; Lego; Shopping Mall; Gateshead; Nun; Piano; Town Square; Elf; Hashtag; Restaurant; Intu Properties	Ai Otsuka; Love It; English Language Films; Music; Performance; Luck; Superstition; 2005 Singles; Christmas; Town; American Films; Russia; Washington D.C; 2006 Albums; Newcastle-upon-Tyne

7.8.3 Comparison of Results

Many of the concepts identified using both Alchemy API and Text Razor, detailed in Table 7.8, can be considered as vague and generic. While such terms will allow for comparisons to be made across a range of domains, they are perhaps not ideally suited for the creation of account-specific, focused engagement profiles. For example, the results generated using these tools highlight 'Metrocentre', 'Shopping Mall', and 'Gateshead' - these are of little use in relation to the Intu Metrocentre, a shopping mall based in Gateshead. While manual analyses may be more time consuming, manual coding conducted by individuals with domain-specific knowledge is likely to produce far richer results, which will lead to more effective profiling of engagement, and therefore facilitate the development of more effective social media strategies.

7.9 Evaluation of Approach, and Implications for Future Use

In the study presented in this chapter, a method for creating profiles of social media users' engagement with content has been proposed and demonstrated. Using this approach, two further aims have been addressed: first, to explore whether users' engagement behaviour and preferences remain consistent (and can therefore be modelled and, to some extent, predicted) over time; second, to investigate whether communities of users that are determined using social network analysis techniques demonstrate similar engagement behaviours. In doing so, this study has addressed the fourth overall objective of this thesis. These are now explored in turn, before the wider implications of adopting such an approach are discussed.

7.9.1 Do Engagement Behaviours Vary Over Time?

In this chapter, a method for developing and implementing engagement profiles for users, and groups of users has been proposed and implemented. Such profiles, however, will only be valid, and therefore useful, if behaviours remain somewhat consistent, and as such the profiles can be used to predict engagement behaviours in the near future. In summary, the results from this study demonstrate that, for the majority of users that have engaged multiple times with the accounts being studied, engagement preferences remain consistent, and therefore can be predicted with some degree of accuracy. In doing so, this has developed on the previously identified gaps in prior understanding, including the need for more longitudinal studies, understanding how users behave online, and how these behaviours may be modelled and predicted.

From the sample of users in this study, totalling over 18,000 followers from two separate retail location Twitter accounts, engagement profiles have differed as little as 5.56% of the maximum possible variation, which takes into account all of the content labels assigned during the qualitative analyses. This approach, however, treats each identified type of content equally, which organisations or individuals are unlikely to do. Organisations are likely to prioritise a subset of content types, and users are likely to have their own individual priorities, shaped by their own experiences (both online and offline), their motivations for making use of social media, and how they manage and present their online persona.

A second means of assessing variation in engagement behaviours is to consider the main types of content with which users do (or do not) engage, which have been termed as their *general engagement behaviours*. These are likely to demonstrate the main content types that users prefer to engage with, or will avoid engaging with, and as such may be of more use than considering each and every content type in combination.

When considering these general behaviours of each user, a total of 45.17% of users, across both accounts, that are present in both the original and validation datasets, maintained identical general engagement behaviours from the original dataset into the validation dataset. A further 33.63% of users maintained the same behaviours, with the addition of another single behaviour to their engagement profile. Combined, this represents a total of 78.80% of engaging users whose original engagement behaviours are still present in the validation dataset.

From the results of this study, based on the data that has been collected and evaluated during this process, it becomes apparent that, when based on sufficient data, there are no major changes to the majority of engaging users' engagement behaviours. As such, these profiles can be used to predict how specific social media users are likely to engage with content over a prolonged period of time. This method of demonstrating, and predicting, user engagement behaviours builds on previous work [171], which highlighted that there is still a need for a greater understanding of the predictability of online sharing behaviours. Not only this, but it also moves beyond grouping users into 'engagers' and 'non-engagers', by demonstrating how users can be grouped based on what types of content they are likely to engage with. Through the research demonstrated in this chapter, this thesis contributes towards this identified need for continued investigation into sharing behaviours, but also highlights that this area is both complex and requiring further investigation.

7.9.2 Does SNA Predict Content Engagement?

Studies presented in previous chapters, such as Chapter 6, have indicated that a user's positioning within a network can be used (to some extent) to predict and model their behaviour - in this case, when they may be at risk of leaving a network, or unfollowing a particular account. Literature discussed in previous chapters has outlined that there are many ways of attempting to understand an audience, such as identifying communities of users within a network. One aim of the study presented here was to evaluate whether those identified as being in a 'community' - i.e. more densely connected to each other than others outside of their community - may engage with similar types of content in a similar manner.

In order to assess this, based on the results from this study, the composition of each of the engagement clusters was evaluated, based on the constituent members' social graph community, as can be seen in Tables 7.6 and 7.7. As discussed previously, these results indicate that, in general, there is no clear relationship between the social graph communities in which individual users are placed, and the engagement clusters that they are assigned to. This would suggest that, while users may form tighter bonds with those that are similar to themselves in some way (based on the principle of homophily), this does not necessarily result in them adopting similar engagement behaviours and patterns. Engagement with online content is a personal choice, with many contributing factors (both conscious and subconscious) at play. While homophily plays a part in who users may be connected to online, this is likely to be based on a narrow set of characteristics. Many such characteristics make up who these individual users are, with individuals choosing, perhaps, to behave online in a particular manner. Existing work has already touched on this, when considering topics such as context collapse, managing multiple audiences, and reputation management.

This has implications for the use of such techniques in the creation and development of social media strategies, and the creation of social media content. While knowing the nature of an account's follower network has its advantages, it cannot be relied upon entirely when developing such approaches and strategies.

7.9.3 Implications for Use and Future Work

In this section, wider implications of this approach are reflected on, focusing primarily on its use in organisational contexts - including the development of social media strategies.

7.9.3.1 Use in Other Contexts

It should be noted that, while the approach taken in this study has focused on its use with Twitter-based data for retail locations, the approach is applicable in many other contexts, and can be easily adapted for use with other (or multiple) data sources. Another context in which this technique can be used, which is outside the scope of the studies presented here, is in the study of cross-platform user engagement behaviours. The techniques demonstrated here could be adapted for use across a range of social media platforms, including those that focus on visual content, in an effort to better understand how users may engage with the same organisation across various platforms, and also how a single organisation's audiences may change depending on the platform being studied. This is particularly pertinent, given the increasing use of many social media platforms [160, 161] and other studies [135] which highlight the need to study user behaviours across multiple platforms, and the potential benefits that may arise from doing so.

7.9.3.2 Supplementing Traditional SNA Techniques

The results of the study presented in this chapter indicate that SNA techniques (such as identifying communities within a network), if used in isolation, may not be entirely effective in the development of social media strategies. Supplementing the information required to make social graphs with, for example, the similarities between users in terms of their engagement behaviours, perhaps indicated by the strength of ties between those two users, could lead to the generation of more useful profile data - identifying groups of users, and the content types with which they are likely to engage.

7.9.3.3 Developing Targeted Content

Combining the profiling techniques that have been demonstrated in this chapter, with techniques for identifying accounts that might be considered influential [113], or bridge accounts [121] within the social graph, would not only identify users to develop content for, but also what types of content they are most likely to engage with, and thus propagate the information further, an act which has already been identified as desirable for those managing social media accounts.

On a broader scale, content could also be created for large clusters of followers who share common engagement preferences. If this analysis is conducted, and the results demonstrate a cluster of 40% of an organisation's engaging audience prefer one type of content, then it may be beneficial for the organisation to consider prioritising that type of content. This, of course, will depend on a number of factors - including the goals of the organisation, the cost of altering their social media content strategies, and the level of reward for encouraging this group of users to engage with content (this may take the form of the number of their followers, for example).

7.9.3.4 Refining Social Media Strategies Over Time

As previously noted, the engagement profiles described in this chapter are based on two separate datasets, collected over the course of an 18-moth period. As would be expected, the accuracy of these profiles increases as more data is available.

Over time, as more data becomes available, such engagement profiles can be updated, thus increasing their accuracy, and allowing for the creation of engagement profiles for new engagers. Adopting such an approach would allow for profiles to be updated on a near-continual basis, with approaches to delivering social media content evolving as the audience, and information known about the audience, continues to evolve.

As discussed in Chapter 3, organisations need to remain adaptable in a market that continues to change and develop. This development of content strategies would therefore, potentially, encourage higher levels of engagement from organisations' online followers, improving the use of social media, propagating an organisation's message further, which has been shown to increase both brand awareness [4], and ultimately brand loyalty [98].

7.9.3.5 Practicalities of Implementation

As presented here, this study involved a large scale post-hoc analysis of Twitter data; this approach could be adapted by organisations to involve the labelling of content at the point of creation, thus reducing the time required to implement this approach. In doing so, the domain-specific knowledge of those closely involved with the organisation and their social media presence would still be maintained. Balancing the benefits of the richness of data produced through this approach, and the potential costs (both temporal and financial), will be an important consideration for organisations.

7.10 Summary

In this chapter, a technique for developing engagement profiles for social media users has been presented and evaluated. As part of this process, engagement profiles were clustered into groups of users that had similar engagement profiles - demonstrating that they have similar preferences as to the types of content with which they are likely to engage. Such an approach relates back to the market segmentation techniques discussed in Chapter 3, instead of segmenting the 'market' based on a particular demographic, these are based on observed online engagement behaviours.

This work is also related to, and builds on, the findings from previous chapters, which indicated differing content strategies engagement behaviours from online audiences (Chapter 5), and that a user's positioning within a network can be used to predict, to some extent, their behaviour (Chapter 6).

As well as demonstrating a technique for developing user engagement profiles, and then clustering these into groups of users with similar demonstrated behaviours, the results of this study have also demonstrated that continued engagement behaviours can be predicted over time. This will be built on in the following chapter, when combined with the findings presented in Chapter 6, the role of prior engagement in predicting and modelling the potential churn of social media followers will be investigated.

Chapter 8

Social Media Engagement and Network Churn

8.1 Introduction

In previous chapters, the use of social network analysis techniques to predict and model follower churn has been explored. These chapters have shown that the position of a user within a social network can be used, to some extent, to model and predict if the user is likely to remain a follower of specified account, or whether they are at risk of 'churning'. Further to this, the modelling and predictability of user engagement with social media content has been demonstrated; the study presented in Chapter 7 shows that, on an individual level, engagement with social media content can be predicted, as the type of content that individuals generally engage with remains relatively consistent over time.

In this chapter, a short study is presented which considers the use of engagement-related statistics in the modelling and predicting of follower churn on online social media. Literature discussed in previous chapters has found that not all social media users will engage equally, or even engage at all. As such, it is unlikely that engagement can be used accurately as a sole predictor for follower churn, but will be useful as a supplementary predictor in addition to those explored in previous chapters, such as Chapter 6.

In this study, various hypotheses are posited, and tested using logistic regression, as in Chapter 6. By comparing the results from this study, with those from the previous study, the efficacy of engagement-related statistics in predicting and modelling social media follower churn can be demonstrated.

The remainder of this chapter is structured as follows. First, a summary of the data used, the various hypothesis and an overview of the analysis methods is provided. Following this, the results from these analyses are outlined, and discussed. Wider implications of these findings are then discussed, before the chapter and its contribution to the thesis overall is summarised.

8.2 Data Collection & Analysis

In this section, various hypotheses relating to the use of engagement statistics in understanding and predicting social media follower churn are stated. Following this the process of collecting and analysing the data to address these hypotheses is described.

8.2.1 Hypotheses

In relation to how engagement-related statistics can be used in the modelling of potential follower churn, the following hypotheses are posited:

H1. When used in isolation, individual engagement metrics (within a single time-frame) will be effective in the modelling of social media follower churn.

H2. When used in isolation, individual engagement metrics (measured to date) will be effective in the modelling of social media follower churn.

H3. The use of engagement metrics as a supplementary predictor, combined with social network analysis metrics, will improve the overall model when modelling social media follower churn.

In each instance, the appropriate null hypothesis will be tested directly, with the alternative hypothesis accepted where the null hypothesis is rejected.

8.2.2 Data Collection & Analysis

As outlined previously, the study presented in this chapter requires the use of engagement-related statistics, and will combine and compare results with the study presented in Chapter 6. In order to aid in this comparison, the same six retail locations were selected to act as case studies, as were used in Chapter 6: Regent Street, London; Carnaby London; Seven Dials, London; Oxford Street, London; Covent Garden, London; and London West End. Details of these locations are provided in Appendix B.

First, it was necessary to determine the number of times each follower has engaged with (i.e. retweeted) a post from the six locations, along with the time the original post was created - thus allowing for each user's engagement to be related to the 'snapshot' in which the engagement occurred. This can then be related to whether or not the user remained a follower, or stopped following the account within that snapshot.

For each follower in each snapshot, the following fields were added to the data used in the previous study:

- Number of retweets - The number of of times the user had retweeted one of the account's tweets during that snapshot.
- Retweeted - A simple boolean field, which indicates whether or not the user has retweeted the account during that snapshot.
- Number of retweets to date - The number of times the user had retweeted one of the account's tweets to date, during the entire time period being analysed.

- Retweeted to date - A simple boolean field, which indicates whether or not the user has retweeted the account to date, during the entire time period being analysed.

Using the data in this way allows for the hypotheses to be tested, considering the use of short-term and long-term engagement in predicting follower churn, and the use of these statistics to supplement the models used in the study presented in Chapter 6.

In a similar manner to the study presented in Chapter 6, each hypothesis can be tested through the use of logistic regression, using the listed variables as independent variables, and whether the user remains in the network as the dependent variable.

8.3 Results

In this section, results relating to each of the hypotheses will be explored and discussed in turn. First, the use of short-term engagement statistics, then the user of longer-term engagement statistics as predictors in social media follower churn. Following this, the use of these statistics as a means of supplementing and improving the models tested in Chapter 6 will be explored.

8.3.1 Short-Term Engagement as a Predictor of Churn

In this study, hypothesis 1 states that "*when used in isolation, individual engagement metrics (within a single time-frame) will be effective in the modelling of social media follower churn*". Accordingly, the null hypothesis being tested is that "*when used in isolation, there is no statistically significant relationship between an individual's engagement behaviour and them remaining as part of the network*". In order to test this hypothesis, two of the previously listed engagement metrics will be used in logistic regression models to test their significance (when used in isolation) in predicting social media follower churn. As in the previous study, the significance of these metrics in predicting follower churn will be summarised.

Number of Retweets

This metric indicates the number of times that a user has retweeted the account during each snapshot. As such, this can give an indication of the usefulness of short-term engagement metrics in predicting and modelling social media follower churn. A summary of the statistical significance of this metric in modelling social media follower churn is summarised in Table 8.1. As can be seen from these results, there are very few instances where the results of the logistic regression would be considered statistically significant (at $p<.050$), with two retail locations having no statistically significant results at all. As such, for this specific metric, the null hypothesis is supported by the data.

Retweeted

Another of the short-term engagement statistics (listed previously) is 'retweeted' - which indicates whether or not the user has engaged within the snapshot being analysed. This is in contrast to the previously tested variable, which demonstrated the number of times the user had engaged. A summary of the statistical significance of this variable in predicting social media follower churn can be seen

Table 8.1: Hypothesis 1 - Number of retweets. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	Mean p Value
Regent Street	1	2	2	.487
Carnaby London	1	1	1	.695
Seven Dials	0	2	2	.701
Oxford Street	0	0	0	.841
Covent Garden	0	0	0	.811
London West End	1	1	1	.701
Total	3	5	6	.706

in Table 8.2. The results for this variable are somewhat more significant than the previously tested variable, with 18.52% of snapshots demonstrating a statistically significant result (p<.050). However, there are no statistically significant results for two of the retail locations, while just over half of the snapshots for London West End have values of p<.050. Again, for this metric, the null hypothesis is accepted due to a lack of consistent statistically significant results.

Table 8.2: Hypothesis 1 - Retweeted. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	Mean p Value
Regent Street	1	2	2	.662
Carnaby London	1	2	2	.605
Seven Dials	1	1	1	.556
Oxford Street	0	0	0	.849
Covent Garden	0	0	1	.765
London West End	2	5	6	.123
Total	5	10	12	.593

Summary

The results summarised in Table 8.1 and Table 8.2 demonstrate that there is no consistency in the statistical significance when using these two variables in modelling and predicting social media follower churn. As such, the null hypothesis is accepted - the use of short-term engagement metrics (in isolation) are not effective in predicting and modelling social media follower churn.

8.3.2 Long-Term Engagement as a Predictor of Churn

In the previous section, short-term engagement statistics - based on observations within a single 'snapshot' - were tested. In this section, hypothesis 2, which considers longer-term engagement statistics will be tested. Hypothesis 2, for this study, is "*when used in isolation, individual engagement metrics (measured to date) will be effective in the modelling of social media follower churn*". As such, the null hypothesis being tested is: "*when used in isolation, there is no statistically significant relationship between an individual's long-term engagement behaviour and them remaining as part of the network*". Two variables will be analysed: first, the number of times each follower has retweeted (to date) posts sent from the retail location, and second, whether the user has retweeted the account at all, to date. The data used for these variables is based on the time-frame analysed over all of the collected snapshots; if implemented by organisations, this could obviously be based on a longer time period, and could be continually updated as more data becomes available.

Retweet Count to Date

This metric indicates the number of times, to the end of the snapshot, that the follower has been observed retweeting the retail location. This metric accumulates over time; a follower that retweeted the account twice in the first snapshot, and a further three times in the second snapshot will therefore have a 'Retweet Count to Date' value of 5 for the second snapshot. This metric was used as the independent variable in a series of logistic regressions. The statistical significance of these results is summarised in Table 8.3. As these results demonstrate, there are very few data points that have statistical significance when used to model and predict social media follower churn. Three of the six analysed locations have no statistically significant results (p<.050), while the remaining three locations have only a single snapshot that demonstrates a statistically significant relationship. As such, for this specific variable, the null hypothesis would be accepted.

Table 8.3: Hypothesis 2 - Retweet count to date. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	Mean p Value
Regent Street	0	1	1	.439
Carnaby London	0	0	0	.474
Seven Dials	0	1	1	.667
Oxford Street	0	0	1	.788
Covent Garden	0	0	0	.578
London West End	0	1	1	.471
Total	0	3	4	.570

Retweeted to Date

The second variable, 'Retweeted to Date' is used to indicate whether or not the follower has engaged (i.e. retweeted) at any point previously. This contrasts with the previous variable, which indicates the number of times they have retweeted to date. Again, this was used as the independent variable in a series of logistic regressions, for each snapshot of each retail location. A summary of the statistical significance of this variable in predicting and modelling social media follower churn can be seen in Table 8.4. Again, these results present a mixed picture. London West End demonstrates 66.66% of results that could be considered statistically significant (p<.050) while many other locations demonstrate a much lower number of statistically significant results.

Table 8.4: Hypothesis 2 - Retweeted to date. Summary of statistical significance of results from logistic regression testing.

Location	p<.001	p<.050	p<.100	Mean p Value
Regent Street	0	1	1	.553
Carnaby London	1	1	1	.403
Seven Dials	0	2	2	.574
Oxford Street	0	2	2	.840
Covent Garden	0	0	1	.461
London West End	4	6	6	.150
Total	5	12	13	.497

Summary

From the results summarised in Table 8.3 and Table 8.4, it is evident that, again, there is little consistency in the statistical significance of results when using engagement-related statistics as the sole indicator in an account follower's likelihood to churn. The results suggest that, in general, whether or not a user has engaged at some point is a more reliable predictor (albeit, slightly) than the overall number of times that they have engaged. However, the results for both of the tested variables can be judged to be ineffective, in isolation, when predicting social media follower churn. As such, the null hypothesis is accepted - when used in isolation, engagement-related statistics are ineffective in predicting and modelling social media follower churn.

8.3.3 Engagement as a Supplementary Variable in Predicting Churn

Previously in this chapter, two hypotheses have been tested, which relate to the use of engagement-related statistics as ineffective sole indicators or predictors of social media follower churn. In both contexts, the null hypotheses were accepted - in that both sets of variables were ineffective in modelling social media follower churn. Here, hypothesis 3 is posited: "*the use of engagement metrics as a supplementary predictor, combined with social network analysis metrics, will improve the overall model when predicting social media follower churn*". Therefore, the null hypothesis being tested is: "*the use of engagement metrics, combined with social network analysis metrics will not improve the overall model when predicting social media follower churn*". In order to test this hypothesis, each of the four variables previously tested will be added to the proposed model in Chapter 6, and the various 'goodness of fit' measures calculated. This will allow these models, now including engagement-related statistics, to be compared to the results recorded in the previous study.

In order to assess how effective each of the following variables are, as supplementary predictors, in modelling and predicting social media follower churn, they were added to the logistic regression model used in Chapter 6. In the previous study, changes in eigenvector centrality, and changes in the outdegree (i.e. the number of 'in network' accounts that the user follows) were used in logistic regression models to determine their use in predicting social media follower churn. A summary of these findings can be seen in Table 6.22 (see Chapter 6), with results for each retail location detailed in Tables C.31 - C.36 (see Appendix C). In the sub-sections that follow, each variable will be tested and discussed in turn.

Retweet Count

This variable relates to the number of times each follower has retweeted a tweet from the analysed account, during each snapshot. Results tables for each location can be found in Appendix D, while an overall summary of these results can be seen in Table 8.5. The results for a logistic regression model consisting of the retweet count, as well as the variables from the previous study (i.e. changes in eigenvector centrality, and changes in the outdegree of each follower) are focused on, while for the sake of comparison, the Nagelkerke's R^2 values from the model from the prior study is also included.

What the results demonstrate is that taking into account this variable improves the fit of the model for each of the retail locations (albeit a small improvement), with the R^2 values increasing somewhere in the region of 0.0007, representing an overall increase in goodness-of-fit of around 0.07%. This would suggest that, while playing some part in predicting and modelling social media follower churn, it is not the most dominant factor in individuals remaining a follower of the social media accounts being analysed in this study.

Table 8.5: Summary of logistic regression results, and comparison to previous model results for engagement & follower churn hypothesis 3 - Retweet count.

	Current Model				Prior Model
Location	Null. Dev	Residual Dev.	AIC	R^2	R^2
Regent Street	1499.40	1482.20	1490.20	0.284632	0.283762
Carnaby London	2399.46	2343.00	2351.05	0.228810	0.227660
Seven Dials	1100.28	1082.26	1090.26	0.319238	0.318300
Oxford Street	1722.74	1636.73	1644.73	0.273157	0.272736
Covent Garden	3195.39	3037.04	3045.04	0.251481	0.250637
London West End	3850.75	3294.95	3302.95	0.286239	0.286201
Mean	2294.67	2146.03	2154.04	0.273926	0.273216

Retweeted

Whereas the previous variable considered the number of times a user has retweeted the accounts being studied, this is a binary variable, which simply represents whether or not a user has retweeted the account during the snapshot. Rather than focus on a scale of engagement, this merely considers the act of engaging to any degree as a predictor in modelling social media follower churn, in conjunction with the two previously selected predictor variables from social network analyses. A summary of the results is presented in Table 8.6, while the results for each snapshot for each retail location can be seen in Appendix D.

Again, the results generated with the inclusion of the additional engagement variable are a slight improvement than the models without the engagement metric, indeed there is a very similar level of increase. This would suggest that there is some limitation on how much the short-term engagement of an individual user can be relied upon when predicting their potential future churn behaviour.

Table 8.6: Summary of logistic regression results, and comparison to previous model results for engagement & follower churn hypothesis 3 - Retweeted.

Location	Current Model Null. Dev	Residual Dev.	AIC	R^2	Prior Model R^2
Regent Street	1499.40	1482.80	1490.80	0.284434	0.283762
Carnaby London	2399.46	2342.84	2350.84	0.228844	0.227660
Seven Dials	1100.28	1081.85	1089.85	0.319574	0.318300
Oxford Street	1722.74	1636.73	1644.73	0.273157	0.272736
Covent Garden	3195.39	2974.24	3044.74	0.251540	0.250637
London West End	3850.75	3293.78	3301.78	0.286570	0.286201
Mean	2294.67	2135.37	2153.96	0.274020	0.273216

Retweet Count to Date

Whilst the previous two tested variables have considered short-term engagement behaviour - i.e. the presence, and level, of engagement within a single snapshot of collected data, this variable considers the number of times, to date, that the user has retweeted (i.e. engaged) with the account being analysed. While this variable is based on information collected within a fixed time-frame, this could of course be continually updated if implemented in an organisational context.

A summary of the results for this model can be seen in Table 8.7, with individual tables for each location shown in Appendix D. Again, these results demonstrate that marginal gains in predicting social media follower churn can be made by considering engagement levels over a longer period of time.

Table 8.7: Summary of logistic regression results, and comparison to previous model results for engagement & follower churn hypothesis 3 - Retweet count to date.

Location	Current Model Null. Dev	Residual Dev.	AIC	R^2	Prior Model R^2
Regent Street	1499.40	1483.09	1491.09	0.284345	0.283762
Carnaby London	2399.46	2343.96	2351.96	0.228455	0.227660
Seven Dials	1100.28	1082.53	1090.52	0.319092	0.318300
Oxford Street	1722.74	1633.43	1641.43	0.275295	0.272736
Covent Garden	3195.39	3038.86	3046.86	0.250729	0.250637
London West End	3850.75	3294.08	3302.08	0.286482	0.286201
Mean	2294.67	2145.99	2153.99	0.274066	0.273216

Retweeted to Date

In a similar vein to the previous variable, this models whether or not a user has retweeted the account being analysed at any prior point in the collected data. A summary of the results can be seen in Table 8.8, with results for each location and each snapshot shown in Appendix D. Again, this shows only a very slight improvement on the existing model, improving the R^2 value by between approximately 0.001 and 0.002, representing a 0.1 to 0.2% increase in the model.

Table 8.8: Summary of logistic regression results, and comparison to previous model results for engagement & follower churn hypothesis 3 - Retweeted to date.

	Current Model				Prior Model
Location	Null. Dev	Residual Dev.	AIC	R^2	R^2
Regent Street	1499.40	1482.06	1490.06	0.284797	0.283762
Carnaby London	2399.46	2344.16	2352.16	0.228372	0.227660
Seven Dials	1100.28	1081.18	1122.93	0.320268	0.318300
Oxford Street	1722.74	1633.54	1641.54	0.275276	0.272736
Covent Garden	3195.39	3038.65	3046.65	0.250816	0.250637
London West End	3850.75	3270.81	3278.81	0.290939	0.286201
Mean	2294.67	2141.77	2155.36	0.275078	0.273216

Summary

Each of the four engagement-related statistics were tested, including them as variables in logistic regression models, along with the two variables identified in a previous study. From these results, it becomes clear that including engagement-related statistics in the models increases their efficacy slightly. Although the inclusion of each variable only has a marginal increase on the efficacy of the model, the largest increase was seen from the inclusion of the 'Retweet to Date' variable. As such, the null hypothesis is rejected, and the alternative hypothesis can be accepted; the inclusion of engagement-related variables, to supplement those ascertained through social network analyses, does improve the efficacy of the model, albeit a marginal increase.

8.4 Discussion of Results

Prior literature has highlighted that not all social media users engage directly with accounts, or indeed produce content, mainly using the platforms as an information consumer, rather than as a producer, which has been demonstrated again within this thesis. Studies explored earlier in this thesis have shown that, to some extent, an account's positioning within a network (determined using social network analysis techniques) can be used to predict and model their likely future churning behaviour. This builds on the motivations identified in prior work (summarised in Table 3.1), which identified an ongoing need to understand the predictability of user behaviour. Further addressing this point, they have also shown that engagement behaviours, in particular the content types with which a user is likely to engage, remain fairly consistent and can be predicted.

In light of this, the effectiveness of engagement-related statistics, when used in isolation, in predicting the potential future churn of social media accounts followers was tested. Two hypotheses were posited; these related to the short-term (hypothesis 1) and longer-term (hypothesis 2) engagement behaviours of the account followers. In both cases, there were very few instances where the results were statistically significant, and no consistency in the results when considered both within and between the different retail locations being analysed. As such, the null hypotheses were accepted in both instances.

Hypothesis 3 looked to assess the improvement of logistic regression models that include engagement-related statistics as well as those from social network analyses. From the results demonstrated in this study, it is clear that the inclusion of such statistics improves the efficacy of these models, but only in a limited fashion. This highlights the previously discussed findings that engagement is only evident in a smaller number of social media users; social media users are often split between those that make use of platforms to produce content, and those that make use of the platforms to almost solely consume content. These findings support the conclusions from the previously discussed literature [60, 73, 200], demonstrating how behaviours can be modelled, and to some extent predicted. These findings also highlight the need for continued research in this area, particularly around the individual motivations for direct engagement with social media content, as well as understanding how 'indirect' or 'offline' forms of engagement may occur [73, 126].

For audiences where the share of engaging users is low, the inclusion of engagement statistics in the churn modelling process is likely to have a very limited impact. However, the efficacy of the models - indicated by measures such as Nagelkerke's R^2 - did improve (albeit only slightly) by the inclusion of these metrics in the logistic regression model. Therefore engagement does have - to some extent - an impact on the retention of users in a social media audience, presenting a further benefit of engagement for organisations.

8.5 Summary

Studies presented in previous chapters have considered the use of social network analysis techniques in predicting and modelling social media follower churn, as well as presenting and evaluating a method for creating engagement profiles for individual social media users. In this chapter, a study was presented that considers the role of engagement in predicting and modelling future churning behaviours of social media users.

Three hypotheses were presented, the first two related to the use of engagement-related metrics as the only predictor in modelling follower churn, the third related to the use of these metrics as supplementary variables in previously tested models. As expected, based on the findings of previous work, engagement behaviour does not function as a sole predictor in modelling social media follower churn. As the share of an audience that actively engages with an account is often low, it leaves a large portion of the audience that have no engagement behaviours to model, and therefore it cannot be used in predicting their churn behaviour. However, when used as a supplementary variable in a logistic regression with metrics determined using social network analysis techniques, the efficacy of the models is improved by a small amount.

This study drew together the findings from studies presented in the previous chapters of this thesis. The initial study demonstrated that retail locations make use of social media in different ways, adopting seemingly different strategies in terms of content that they share, the levels of engagement, and the growth in their online audiences that they receive as a result. The studies following this explored different aspects of these interactions - predicting and modelling audience churn, and modelling engagement with types of content. This study drew these together, exploring how engagement with content might be used, and to what extent, in predicting social media follower churn.

Chapter 9

Discussion

9.1 Introduction

Previous chapters in this thesis have discussed related research and theories, methodologies of data collection and analysis, and described four separate studies into understanding and modelling social media behaviours. Each chapter which presents a study also includes a discussion of the findings of that study, along with the implications of these findings. This chapter presents broader discussions, based on these findings and research contributions. The chapter is structured as follows. First, a brief overview of developments in relevant academic fields since this research was conducted, is provided. Following this, a review of research contributions is provided, to contextualise the remainder of the chapter. Third, the potential implications for practice - the use of the approaches demonstrated in this thesis in an organisational context - is explored in more detail. Potential limitations are then discussed, both in terms of the general approach to social media research, and then the individual studies and approaches are discussed. The chapter concludes with a discussion of implications for future research and further development, including areas that require further ethical consideration.

9.2 Recent Developments in Academic Literature

The work presented in this thesis has been conducted over a number of years, beginning in 2013, with the majority of data collection concluded during 2017. As such, the contributions made within this thesis should be weighed against the ‘state of the art’ as it was then. In the intervening years since this programme of research was first designed and instigated, there has been progress made in this, and related fields of research, as well as commercial practice, as well as the development and introduction of new social media platforms.

One of the major motivations for the research within this thesis was the need to better understand how individual users engage with social media content, and the ways in which this engagement might be understood and modelled or predicted. Various related literature has been published in this field, some of which is directly related to the methods used within this thesis, while other publications have investigated slightly different aspects of engagement.

For example, Dolan et al [57] also investigate the effect of particular types of content on engagement levels from social media users. While the studies presented within this thesis focus on different types of content in Facebook posts (Chapter 5), and then build engagement profiles based on Twitter content (Chapter 7), this paper considers content in four categories: informational, entertaining, remunerative, and relational. In many ways, these categories are similar to the themes identified within the analysis in Chapter 5, and are again present in the content codes used within Chapter 7.

Other studies have focused on the content of images within social media posts [117], which is one aspect that has not been directly considered in the studies presented within this thesis. However, the methods demonstrated and utilised within Chapter 7 could easily be adapted for use with visual content. Further studies have also considered the format of the post, and the platform on which it is shared [170], as well as the readability of the post itself [157] in relation to the resulting engagement. While the studies in this thesis do not consider the effect of individual platforms on engagement, the method demonstrated in Chapter 7 could be used on a cross-platform basis, and could be used with a variety of content types.

The related research that has been published since this programme of research was devised and implemented demonstrates the continuing interest in this field of work. While the research that has been briefly outlined here is not a direct replica of the studies in this thesis, this literature would have, perhaps, informed and shaped this thesis in a different manner, if it had been available at the time. For example, with more research considering the content of social media posts and the resulting engagement levels, then their grouping of content, or coding schemes, may have been used, or informed the schemes used within this thesis. Further, these developments and moving focus toward visual and cross-platform studies would perhaps have resulted in a keener focus on a cross-platform understanding of social media engagement.

In the next section, the research contributions made within this thesis, particularly in relation to the 'state of the art' when the studies were first devised, are reviewed and discussed.

9.3 Review of Contributions

As outlined in Chapter 1, this thesis presents a number of research contributions, with an overall view to developing a greater understanding of how organisations' social media audiences behave and develop over time. These contributions are reiterated here as a means of contextualising the discussion in the rest of this chapter. Table 9.1 further develops Table 3.1, by demonstrating the links between research motivations, research objectives, the studies presented within this thesis, and the resulting contributions.

First, a technique for profiling the engagement behaviours of social media users has been presented (see Chapter 7). This technique has been demonstrated to model the general engagement behaviours of social media users; further, these behaviours have been shown to remain consistent and predictable, in the majority of cases, over time. While this approach has been demonstrated with data acquired from Twitter, it is, by design, flexible in nature in terms of the types of content it can be used with. As such, this approach can be adapted for use with other platforms, that focus on other types of data, such as visual data from Instagram, for example. This builds upon the identified motivations

Table 9.1: Table demonstrating the relationship between motivations for research, research objectives, thesis studies, and research contributions.

Research Motivations	Research Objectives	Thesis Studies	Contributions
The need for longitudinal studies, as identified by Easley & Kleinberg [60]. Facilitating the further development of organisations moving to a post-modern marketing focus [166]. Understanding what is being shared online [126].	RO1 - To develop an understanding of how retail locations make use of social media, including the range of content being shared and how this may affect the levels of engagement from their social media audiences.	Study One (Chapter 5), Study Four (Chapter 8)	An understanding of the types of content shared online by retail locations, and the correlation between different types of content being shared and the resulting relative levels of engagement. Demonstrates no general consensus across organisations as to what is the 'most engaging' type of content, demonstrating that each audience is different, and therefore there are different preferences and reasons for engaging, requiring further research.
The need for longitudinal studies, as identified by Easley & Kleinberg [60]. A need to understand the reality of underlying networks in such studies, as per Friedkin & Johnsen [73]. Facilitating the further development of organisations moving to a post-modern marketing focus [166]. Further understanding 'churn' in an online social media context. How individuals' position in a network may effect churn.	RO2 - To develop a greater understanding of how online social media audiences may grow (or decline) and develop over time.	Study Two (Chapter 6)	A longitudinal study of the growth of social networks, demonstrating that different networks grow at different rates. This informs and motivates further studies in this thesis, and a need to investigate further in future work.
The relative importance of social network metrics (when used in context), as raised by Wasserman & Faust [200]. A need to understand the reality of underlying networks in such studies, as per Friedkin & Johnsen [73]. Facilitating the further development of organisations moving to a post-modern marketing focus [166].	RO3 - To investigate the extent to which social network analysis techniques and metrics can be used to indicate the likely organic growth or decline of social media audiences.	Study Two (Chapter 6), Study Four (Chapter 8)	Building on contributions of RO2, an understanding of how various SNA metrics can be used as indicators to predict audience churn. Demonstrates that while some metrics are useful (in this context) for modelling churn, other factors still need to be taken into account to improve the models. Again, this motivates future work in this context.

Research Motivations	**Research Objectives**	**Thesis Studies**	**Contributions**
The need for longitudinal studies, as identified by Easley & Kleinberg [60]. Facilitating the further development of organisations moving to a post-modern marketing focus [166]. Further understanding 'churn' in an online social media context. Understanding what is being shared online [126]. Understand the predictability of sharing behaviours, and preferences for sharing [171].	RO4 - To implement and evaluate a method of generating profiles of social media users' engagement with specific content, to enable retail locations to understand the aggregate behaviour of their online audiences	Study Three (Chapter 7)	A means by which engagement with social media content can be profiled. Further, how these profiles can be grouped and understood at scale, and how these groupings compare to more 'traditional' means of grouping members of a network.
The need for longitudinal studies, as identified by Easley & Kleinberg [60]. Facilitating the further development of organisations moving to a post-modern marketing focus [166].	RO5 - To develop an understanding of the role of engagement with social media content in the growth and development of social media audiences.	Study Two (Chapter 6), Study Three (Chapter 7), Study Four (Chapter 8)	A better understanding of the role of engagement levels in predicting social media follower churn, and in turn how this may impact network growth. A better understanding of combining historic engagement levels with social network metrics in modelling and predicting audience churn.

from prior research, including the need to consider behaviour across and between different platforms.

Through clustering such profiles together into groups of users who demonstrate similar engagement patterns and preferences, those utilising such approaches can develop an understanding of their audience, and those within it that are likely to engage with a given content type. These groups have been demonstrated to be different from those communities detected and determined using widely-adopted social network analysis techniques. This demonstrates, for example, that communities of users that are highly connected will not necessarily demonstrate the same engagement behaviours and preferences toward a given type of content. This has implications for how entire networks of social media followers could be divided and grouped in order to develop a greater understanding of the audience - these are explored more fully later in this chapter.

Second, a longitudinal study (motivated by the identified need for longitudinal studies in this area), presented in Chapter 6, considered the notion of 'churn' within an online audience, and how those at risk of churning might be identified through the use of various metrics relating to their position and connectedness within the network of social media followers. This study was based on the findings of the initial study detailed in Chapter 5 which first discussed the growth of social media audiences. Whilst many metrics, in isolation, were demonstrated to be ineffective in modelling and predicting follower churn, a small selection of these metrics were shown to have value in this regard, with a combination of these metrics shown to be the most effective.

The final major contribution of this thesis relates to the role that engagement may play in the growth and development of online audiences. A further study presented later in the thesis (see Chapter 8) considered the use of engagement-related metrics, first in isolation, and then in conjunction with the most effective combination of metrics demonstrated in the prior study. This highlighted that, while encouraging engagement may result in those individuals remaining part of the network, the proportion of followers who actively engage with social media content is too low to make it an accurate predictor when used in isolation. This develops the findings of previous work, which identified that users can often be divided into those that actively engage and those that do not, by demonstrating how this engagement may lead to retaining social media followers. Using such metrics in conjunction with those related to a user's position within a network were shown to have a small increasing effect on the effectiveness of such models. This has potential implications, which are discussed more fully later in this chapter, for those that may rely entirely on engagement levels and metrics as a means of judging the 'success' and potential growth of their online audiences.

9.4 Potential Implications for Practice

In this thesis, multiple methods of developing a greater understanding of an organisation's online audience – and their behaviour – have been demonstrated. These include developing a means by which their engagement with specific types of social media content can be modelled and predicted, and understanding and identifying portions of an online audience that are at risk of 'churning'. Such contributions have potential implications for practice, which are now explored here. The discussion here is presented in two main areas: the need, and benefits, of understanding the behaviour of online social media audiences, and how these approaches could be appropriated and implemented by organisations. As outlined earlier in this thesis, while the context in which these are presented is mainly related to retail organisations, such approaches are flexible in nature and are suitable for use by organisations in a variety of contexts.

9.4.1 Understanding Audience Behaviour

The need to understand one's audience, and the benefits of doing so, is not new, and indeed has not been introduced with the advent of social media. However, the affordances of social media, such as the 'breaking down' of geographical boundaries [174], the ease of joining such platforms, and the immediacy of communication, both exacerbates this need, and dramatically expands the potential audience of such organisations [174].

Social media offers organisations the opportunity to share their message – whatever that may be – with a wide audience, both in terms of number and (potentially) geographical location [174]. For this to be successful, organisations thus require the audience to be numerous (the scale of this will differ, depending on the nature of the organisation and their target market) and engaged. It then follows that organisations would benefit from understanding how to maintain, add to, and encourage engagement from, that audience.

Recently published research supports this continuing need to further understand audience behaviours. This includes how individuals may be modelled [44], the characteristics of influential people [62], and how information and knowledge is transferred and exchanged in networks [210]. While the work presented here goes some way to contributing to this understanding, further work is required as platforms continue to develop, and so does user behaviour in response to these changes and developments.

9.4.2 Potential Implementation by Organisations

The continued interest by organisations to make use of their social media presence is evident by recent commercial articles and publications, covering a range of relevant topics. These topics, such as correctly understanding engagement rates [206], making use of a range of social media platforms [94], and changes in audience age range and behaviours [10] show the development of how organisations are now looking to utilise social media. Through the findings of these studies, and the implementation of the methods demonstrated here, organisations can benefit from a greater understanding of recent engagement behaviours and changes in their audiences.

As discussed previously, organisations will benefit from a greater understanding of their audience – and its behaviour – as a means of maintaining, encouraging the growth of, and encouraging engagement from, the audience. Such an understanding, however, will be of no value if organisations are not willing, or able, to implement any suggested or required changes to their operating processes, such as the creation of social media content. Here, a suggested implementation of these techniques within an organisation is discussed. Again, while this focuses primarily on organisations within a retail context, such approaches are valid in many other contexts. While organisations are likely to have a presence on various social media platforms, and the use of these platforms should be planned and coordinated, the affordances of these platforms are distinct, and as such lead to differing approaches when understanding audiences, and acting on this information.

For organisations making use of social media platforms such as Facebook and Twitter, there are a number of potential goals or needs that the organisation may have (some of these may be known to the organisation, whilst others may not). These include: a need to understand their audience, including 'key' or influential individuals [60, 179]; a need to maintain their current audience, whilst also encouraging the growth of that audience; the need to maximise dissemination of their posts; a need to encourage engagement, as means of increasing the previously mentioned dissemination of their posts. Each of these could be achieved through the implementation of the various techniques explored within this thesis.

First, organisations can look to develop a greater understanding of the structure of their audience. Social network analysis techniques (such as detecting communities [78, 144]) may not indicate groups of users that are likely to engage with the same content. However, they can be used to identify the structural nature of the audience, which includes the identification of key, or important, individual accounts [60, 179] within that audience. Such individuals may be identified as such if, for example, they have a high number of followers, or are positioned in a key position of the audience – such as being a 'bridge' account between two subsets of communities within the audience. Being able to identify such accounts may inform how the organisation choose to produce social media content in the future, depending on their goals. Retaining followers that themselves have a high number of followers may be beneficial as they may then encourage their own followers to follower the organisation's account. Further, encouraging these accounts to engage (i.e. retweet) will then disseminate the original post to a much wider audience than with an account with a small number of followers.

Second, organisations can look to develop a greater understanding of how their current audience engages (or does not engage) with the content currently produced and shared from the organisation's account. This builds on prior work, which identified engagers and 'non-engagers', and the ongoing need to develop a better understanding of how individuals engage, and the motivations for doing so. This could be done using the technique demonstrated and evaluated in Chapter 7, and can be used with any data source where engagement actions are accessible – such as retweets or likes on Twitter, and shares and comments on Facebook, for example. This will provide two types of information to organisations, and those managing with organisation's social media presence. It indicates what types of content receive the most (and least) engagement, and thus where the organisation may wish to focus and refine its content. It also indicates which users are frequently engaging, and with what types of content. This can then be combined with information such as which users are 'key' or 'important' [72, 73], to aid the organisation in focusing its social media content strategies. For example, if a very popular follower engages with a specific type of content, then account managers may wish to produce more content of that type. In contrast to this, if 80% of the content shared by an organisation is of one particular type, and receives very little engagement, then the organisation may wish to refocus its efforts on other types of content, in order to encourage higher levels of engagement. While Chapter 7 demonstrates this approach with a large-scale ad-hoc analysis of content types, social media posts can be labelled at the point of creation, thus reducing the perceived workload of using this approach.

The third way in which such approaches can be used by organisations is to understand the volatility of their online social media audience. As demonstrated in Chapter 6, social network metrics can be used to both determine an account's position within a network, but also (over time) the likelihood of that account 'churning' - in this context, stopping following the organisation's social media account. This approach can be used in situations where relational data is available – such as Twitter – indicating relationships between two given accounts on that platform. By recording and analysing the state of their follower network at regular intervals, accounts can be highlighted that are perceived to be at risk of churning, and then a decision model followed which evaluates the benefits of retaining that follower against the potential or perceived cost of acting to attempt to retain the follower. For example, if a 'high-value' account (potentially identified from the previously explored techniques) is seen to be at risk of churning, then actions can be taken to attempt to retain that follower. This could include creating 'targeted' content (though not necessarily directed at the account) with which they are likely to engage - again, these types of content can be identified from the previously discussed technique of engagement profiling.

Though explained here as a combined set of techniques and processes, such approaches can be used in isolation, and supplemented with other approaches and techniques that the organisations may already have in place. The changing nature and affordances of existing platforms, and the introduction of new platforms, requires a flexible approach to managing and refining social media presence, and as such these techniques and approaches are intended to be applicable in a variety of contexts, and with a range of different types of content.

The nature of each individual organisation may also impact how they look to make use of such approaches. A small organisation making use of social media for the first time, for example, may look to get an understanding of what types of content result in the highest levels of engagement, without looking to understand this engagement on a per-follower basis. This may contrast to larger organisations, with more resources assigned to managing their social media presence, that may wish to develop a more comprehensive understanding of their social media audiences and have the opportunity and resource to refine their online presence in a more comprehensive manner.

9.5 Potential Limitations

The studies within this thesis have considered various aspects of understanding social media audiences, such as their engagement behaviour and the volatility of online audiences and understanding when individual users are at risk of 'unfollowing' a social media account. Though such techniques and approaches aid in developing this understanding, there are potential limitations, some of which may be addressed in future work in this, and related, fields of research. Here, these limitations are discussed in two parts. First, more general limitations to the overall approach are discussed. Second, these are related to each of the studies presented in this thesis, including a discussion of the effects of these limitations on the results and findings of each study.

9.5.1 Limitations of Overall Approach

First, the studies presented in this thesis have focused on a single social media platform within each study (primarily Twitter), and only Twitter and Facebook have been considered across the entire thesis. While organisations may make most use of these platforms, many have a presence on other platforms that have not been considered. However, many of the approaches described within this thesis can be used with other data types and sources, such as social media platforms that focus on visual content, such as Instagram, Pinterest, and YouTube. Recent research [105] has considered engagement with video content, specifically on YouTube. As has already been discussed, organisations are likely to make use of multiple platforms simultaneously, and as such there is a growing need to understand how the audience makes use of, and engages across, multiple platforms.

Further, literature explored earlier in this thesis highlighted that users can often be grouped into two main categories – those that are prone to engaging with, and producing, social media content (i.e. producers), and those that make use of social media in a more passive manner (i.e. consumers) [142]. By definition, the engagement preferences of individuals can only be modelled and predicted if the users have already engaged with existing content. In the studies presented in the previous chapters, the focus has been on engagement with content produced and shared from the specific organisation's social media account. In the case of 'non-engagers', it may be possible and beneficial to investigate further to ascertain if they engage with other accounts, perhaps even those in the same or similar market to where the organisation operates. If so, it may be possible to model their engagement preferences on content from other accounts. In such cases, the demonstrated techniques can be adapted and used with data from other sources (i.e. relating to other accounts). This once again highlights the need for a greater understanding of the motivations behind engagement on social media, and is reflected in the more recent related literature. While this thesis has contributed to developing an understanding of how individuals engage and with what content they engage, the underlying motivations of engagement (which are likely to be numerous, complex, and entirely dependent on the individual user) remain an area requiring further investigation.

Third, such approaches and techniques require a level of 'buy-in' from organisations. Developing, and utilising, a greater understanding of an online audience has costs associated with it – likely both temporal and financial. However, such costs can be minimised, and the potential returns of these approaches understood and valued if organisations adopt them into their existing working practices. While such techniques and approaches allow for a greater understanding of an audience, the value is generated from acting on this new understanding, to benefit and develop the social media presence of an organisation.

Fourth, and finally, social media research (at present) often relies on the use of each platform's API. As such, researchers and organisations alike are affected by any changes and restrictions applied to these APIs. This can have implications for ongoing studies, or business processes, particularly when access is not only altered, but removed entirely. In April 2018, for example, Instagram closed their public API, overnight, and without warning. Although upcoming changes had been announced, these were scheduled for much later in 2018. As a result, researchers and organisations lost access to this source of data immediately. This is, however, a rare case. Although platforms continue to adjust and alter how data is accessed, and how much data is accessible, many of these changes are planned and announced in advance, with academic researchers and organisations within a genuine business use-case given access to the data. Data collected for the studies presented in this thesis, for example, is still available from both Facebook and Twitter, although the methods of collecting such data have changed. While this does not necessarily prevent or restrict research in this area, it is still a potential limitation - researchers and those working in and for organisations will need to be aware of upcoming changes to API access, and respond accordingly.

9.5.2 Limitations of Individual Studies

In this subsection, the limitations of each of the four presented studies are presented and discussed, including the impact that these limitations may have on the findings of each study.

9.5.2.1 Study 1 (Chapter 5)

This study included an analysis, covering a period of 12 months, of Facebook post content shared by retail locations, and the resulting level of engagement, as indicated by the number of likes each post received. Extending such an analysis, covering more data points than a 12-month period would lead to the results of analysis techniques such as Spearman's Rank being more meaningful.

By focusing solely on Facebook accounts, the findings of the study cannot necessarily be extended to the use of the social media platforms by retail locations. Further, the consideration of likes as the sole engagement metric ignores the potential for other forms of engagement - such as comments, and shares. If these were to be included in future work, then the overall engagement levels for different types of content may be different.

Finally, since this study was conducted, the access to Facebook data APIs has been severely restricted. While this may impact the ability of individual researchers to access this data in the same manner, such APIs are still accessible by organisations with legitimate interests in doing so. As such, this should not impact the ability for organisations to understand the content being shared by their competitors, or their own accounts, and the resulting levels of engagement.

9.5.2.2 Study 2 (Chapter 6)

The second study in this thesis, presented in Chapter 6, focuses on the growth of social networks over time, and the use of social network analysis metrics as a means of modelling and predicting follower 'churn' in these networks. While this study covered a total period of 18 months, and consisted of a dataset of millions of data points, there are limitations which impact the findings in various ways.

First, this study focuses on what is a very small subset of potential accounts in this area. While the findings relate to the accounts analysed within this study, further investigation would be required - on a larger dataset, to assess if the findings hold true in a wider context. These findings do, however, indicate that these metrics are useful, to some extent, in modelling and predicting account follower churn.

Second, this study focuses on the retention of individuals already within the network, i.e. those that are already followers of the account being studied. What this study does not do, is take into account how these accounts may be responsible for attracting new followers to the account. Future studies could achieve this, by extending the data collection further to include the network of 'friends of followers' - i.e. those that are one further 'layer' from the account being studied. In a practical context, this would also help organisations understand how specific individuals could be motivated to share content with their own followers, as a means of increasing their follower network. This would further develop such a study to account for network growth in terms of both retaining existing followers, and attracting new followers.

9.5.2.3 Study 3 (Chapter 7)

This study, presented in Chapter 7 of this thesis, presents and evaluates a technique for developing profiles that model the engagement preferences of social media users. This technique is based on the types of content with which these users engage,

The first limitation of this study, and this approach, is that it only considers the Twitter platform, and specifically the retweet, as an indicative means of engagement. While this provides a very narrow focus to the study as it is presented here, the approach to modelling users' engagement behaviours can easily be extended to include other platforms, and other engagement metrics.

Second, and a limitation of many such studies in this field, is that it does not account for engagement with content that may happen 'offline', or in an indirect manner on social media. Designing and conducting such a study, on such a large scale, would be difficult but would give greater insights into how engagement happens, in its various forms, on and off social media.

Third, and related to the previous limitation, is that there has been no direct involvement with the individuals that are represented by these social media accounts. Directly involving individuals in such a study, or studies, would allow previously discussed topics and theories such as self-categorisation theory, and prototypical account followers to be addressed. This would potentially deepen the insights gained from such a study, as to the real motivations behind engaging with specific types of social media content. However, such an approach would not be feasible for every follower of a social media account.

9.5.2.4 Study 4 (Chapter 8)

The fourth and final study, presented in Chapter 8 of this thesis, builds directly on the previous two studies, presented in Chapters 6 and 7. As such, the same limitations apply to the study presented here. These limitations include the narrow focus on Twitter, and specifically the retweet function, as well as limitations such as not taking into account other means of engagement and information diffusion.

9.6 Implications for Future Research

The findings from the studies presented here, and the analysed literature, highlight the need for continuing and further research in various areas; those detailed here include: cross-platform studies and studies which focus on other platforms, further longitudinal studies, a greater understanding of the motivations for online engagement, and various ethical considerations that arise as social media use and technologies continues to develop. Each of these are explored in more detail in the following sections.

9.6.1 Cross-Platform Studies & Studies on Other Platforms

Many studies relating to social media often investigate a single social media platform, or a selection of platforms in isolation. However, both organisations and individual users make use of multiple platforms concurrently, maintaining a presence on each. Each platform has its own affordances, and audiences, and as such both organisations and individuals are likely to use each platform differently, and for different reasons.

It follows, therefore, that future studies need to consider how users make use of a range of social media platforms, in isolation and combination, as well as understanding their behaviours (including engagement) across these platforms. Recent research [85] has reiterated this need to continue considering the interplay between technologies and group dynamics. Users may interact with the same organisations across a range of these platforms, but behave in different ways depending on the platform, its affordances, and their perceived audience on each platform. In the same way that organisations may consider marketing strategies for television, radio, and print media as separate but related, organisations (and therefore related research) should consider social media platforms in a similar way.

9.6.2 Long-Term Studies

While not feasible in the context in which these studies were conducted, a longer-term study in which user behaviour was modelled, social media content specifically designed and shared, and the effects analysed would further validate such approaches to managing social media.

Complementing the studies presented here, research published recently has highlighted not only the need for greater inclusion of social theory when understanding online group behaviours, but also how technologies may impact long-term group (i.e. audience) viability [85]. Further to this, social media platforms, their design, and affordances are often changing. Therefore, such studies, conducted over a prolonged period of time, would be able to assess the effect and impact on user behaviour, including engagement and network development. For example, organisations already modelling user behaviours would be able to see the changes in these behaviours and adjust their social media strategies and content where appropriate.

9.6.3 Understanding and Encouraging Engagement from 'Non-Engagers'

Understanding the motivating factors that encourage individual social media users to actively engage with social media content still remains an area of research that requires further study. Such factors are likely to be numerous, complex, and differ from individual to individual. Burbach et al. [44] for example, highlight the relationship between the personality of individuals and their online behaviour. What also needs to be understood is how (and if) the individual outcomes of passive engagement – such as reading a post – may differ from active engagement, such as reading and then sharing a post. While the latter is perhaps easier to measure and assess quantitatively, passive engagement will have its own outcomes, particularly for the individual. These outcomes may benefit the offline performance of organisations, or may result (through word of mouth, rather than electronic word of mouth, for example) in seemingly unrelated online engagement from other accounts.

9.6.4 Ethical Considerations

The collection, analysis, and use of social media data continues to cause debate and discussion around various aspects of what is, or should be, considered ethical research. Many social media platforms allow access to 'public' data through the use of their own APIs. This data is deemed to be 'public' in that users have not marked their profiles (and therefore their posts, and activities) as private, and the use of their data was made clear in the Terms and Conditions that they accepted on signing up to the platform.

In more traditional research environments, the concept of informed consent dictates that research subjects should be informed, and give their consent, to be involved in a research study. However, with the abundance of data readily available through social media APIs, this becomes both impractical and seemingly unnecessary. Various guidelines [33, 34, 35, 36, 37] highlight the need for maintaining anonymity in these situations, through the use of pseudonymization and obfuscating example post content through merging similar posts and the paraphrasing of posts. Such guidelines are widely adopted and are evident in the majority of recent research based on social media. Other research has also higlighted that, depending on the research context, researchers may actually put themselves in danger if approaching social media users to request permission to include their data within a study's dataset [74].

What remains a somewhat under-studied area of research is the views of those that become 'participants' in social media studies. A recent study [67] has begun to address this issue, with a survey of 368 participants. The views of social media users regarding the use of 'their' data in academic research, without their knowledge, are mixed. While some users are content with their social media activity being used without their knowledge and consent, others suggest a need to be informed – either at the point of data collection, or at the point of conclusion. While the number of participants in that study is low compared to the total number of social media users worldwide, it does highlight the fact that there will be (and are) a range of opinions and points of view on this topic.

Further research in this area could consider the range of views on these matters and, for example, how this might differ in different geographical or cultural contexts. This would help inform the creation and implementation of guidelines that take into account the views of data 'subjects' as well as those conducting the research.

9.7 Summary

This chapter has focused on a discussion of the findings and implications of the research presented in this thesis. Drawing together the findings and contributions from the various studies, this chapter highlighted how such approaches can be implemented by organisations, and how future research might address various points that were raised during these studies. Highlighting the need for continued research in the field of online engagement, this chapter concluded with four broad areas that, it is suggested, still need to be addressed.

Chapter 10

Conclusion

10.1 Introduction

To conclude this thesis, this chapter reviews the initial aim and objectives as outlined at the beginning of this thesis, highlighting how each of these was addressed through the presented studies, and how the findings from these studies not only informed other studies presented here, but also areas that might be addressed in future work.

In earlier chapters, related literature and social network analysis techniques and concepts were explored. These included prior work on identifying demographics and behaviours of social media users [158, 159, 165, 176, 177], as well as organisational use of technology, including online social media and the business benefits of doing so [66, 106, 129, 174]. This highlighted the benefits of encouraging engagement from social media followers, and the ways in which a social network can be modelled and understood.

A total of four studies have been discussed in this thesis. The first considers the use of social media by retail locations, focusing on Facebook as an example platform, with discussion around the types of content being created and shared, and the levels of engagement received as a result. The second study builds on this, and looks at the development of networks of followers surrounding the social media accounts of retail locations, focusing on the use of social network metrics to model and predict when individual users are at risk of 'churning'. In the third study, a method of generating individual engagement profiles for social media users is presented, implemented, and evaluated; this study demonstrates that the general engagement behaviours of social media users often remains fairly consistent, and can therefore be modelled, predicted, and used in the creation of social media strategies. The fourth study considers engagement-related metrics as a means of predicting social media follower churn, and demonstrates that these metrics, in isolation, do not serve as an effective means of doing this; however, when used in conjunction with the models used in the second study, then the efficacy of these models is slightly increased. In each chapter that presents one of these studies, a discussion of the results and their implications was provided. Following these chapters, in Chapter 9, the broader implications of the research presented in this thesis were discussed, including ethical considerations that, in general, remain unaddressed.

10.2 Review of Aims and Objectives

In Chapter 1, the aims and objectives were outlined. The overall aim of the research presented here was to develop a greater understanding of social media users' engagement behaviours, and the effect that this engagement may have on maintaining and growing online social media audiences.

In order to achieve this aim, five distinct objectives were identified. These objectives were based on identified motivations from literature, as initially detailed in Table 3.1. These were then revisited in Chapter 9, Table 9.1, where the motivations, research objectives, studies, and contributions are all outlined and related.

Each of these objectives are now listed in turn, along with an overview of how this objective was addressed within this thesis, and the contributions made as a result.

10.2.1 Objective One

The first objective was: "*to develop an understanding of how retail locations make use of social media, including the range of content being shared and how this may affect the levels of engagement from their social media audiences.*"

This objective was addressed in the study presented in Chapter 5, through the initial study which considered the Facebook accounts of six distinct retail locations. Through an analysis of the types of content shared from these accounts over time, and the levels of engagement received, various results were highlighted.

First, it became apparent that different organisations make use of social media in different ways, with the creation of various types of content. In turn, this resulted in varying levels of engagement from their audience. It was also demonstrated that there is no strong correlation between content type and engagement when considered between organisations, highlighting that there is a need for organisations to understand their own audiences, and how this may impact the levels of engagement received as a result. These findings were used to inform the focus of the studies that followed.

10.2.2 Objective Two

The second objective was: "*to develop a greater understanding of how online social media audiences may grow (or decline) over time*."

This objective was addressed predominantly in the study presented in Chapter 6, through a longitudinal study of the growth of the online audiences of six retail locations' Twitter accounts, as well as through some of the findings from the study presented earlier in Chapter 5 and is linked to objective three. These results showed that the audiences of different organisations' accounts (as might be expected) grow at varying rates. This again highlights the need to develop and understanding of an organisation's audience, the factors that may affect its development, and as a result understand what approaches might be taken to encourage this. This not only informs the rest of the studies presented here, but also has implications for further work in this area of research – including the need for further investigation into the potential motivations for individuals following, and actively engaging, with organisations through social media.

10.2.3 Objective Three

The third identified objective was: "*to investigate the extent to which social network analysis techniques and metrics can be used to indicate the likely organic growth or decline of social media audiences*."

This objective was addressed in the study presented in Chapter 6, utilising the same dataset collected to address the second objective. Through various analyses conducted using regression testing, it was demonstrated that some network metrics are useful in helping to predict and model when social media users may be at risk of 'churning' – that is, in this context, of unfollowing the organisation's account. These results also indicated that while these metrics, which relate to an individual account's positioning within the network, can be used to some extent, they do not account for every influencing factor that may impact churning behaviour, and as such, further investigation and research is required in this area. This both influenced a later study in this thesis, and also highlights that the motivating factors behind the online behaviour of individuals are complex and multi-faceted, and cannot be fully understood, modelled, and predicted using only a narrow set of data and metrics.

10.2.4 Objective Four

Objective four, outlined at the start of this thesis, was: "*to implement and evaluate a method of generating profiles of social media users' engagement with specific content, to enable retail locations to understand the aggregate behaviour of their online audience*."

Chapter 7 presents a study in which this objective was addressed. A process by which engagement profiles can be generated, analysed, and grouped was proposed, demonstrated, and evaluated. This approach is flexible and can therefore be adapted and appropriated for use with a variety of data types, and therefore social media platforms. This study also demonstrated that, broadly speaking, social media users are consistent in the main types of content from a a specific account with which they will engage; as such, the results of these profiles, and the grouping of users with similar behaviours, can be used in the strategic planning of social media content and campaigns, to encourage higher levels of engagement from a known and profiled audience.

10.2.5 Objective Five

The fifth, and final, objective outlined at the beginning of this thesis was "*to develop an understanding of the role of engagement with social media content in the growth and development of social media audiences*."

Addressed in the study presented in Chapter 8, this objective draws together the previous objectives and studies. With engagement behaviours understood and profiled, and the methods by which social network growth may be understood (to some extent) by social network analyses and metrics, this objective sought to understand how engagement behaviours might contribute toward the modelling of potential network growth or decline.

Through an initial investigation, using engagement-related metrics as independent variables in regression models, it was demonstrated that, in isolation, engagement cannot be used as the sole predictor in potential churn behaviour from social network followers. These metrics were then used as supplementary predictors in the models identified and tested in Chapter 6. These analyses demonstrated a marginal increase in the efficacy of these models when engagement was considered alongside the social network analysis metrics previously used.

Whilst these models were improved slightly, it further reiterates that, in general, the portion of an audience that actively engages with a social media account is small. As such, there is a need for future research to pursue two broad areas, which are detailed in the following section.

10.3 Implications for Practice

In the previous chapter of this thesis, practical implications of this research were discussed in Section 9.4).

In terms of practical implications, the research included within this thesis provides a means by which organisations can understand, on a detailed basis, how the content they are producing and sharing impacts the engagement from their social media followers. The discussions included in Chapter 9 also outline how the profiling technique demonstrated in Chapter 7 can be integrated into the operational practices of organisations. This also includes how this may be combined with the findings of the studies in Chapter 6 and Chapter 8, which consider 'churn' within the network of their followers. By including such a technique in their operations, organisations will be able to understand how elements of their social media followers are engaging with their content, and which of their followers may be at risk of 'churning'. This will allow them to make changes to their content creation strategies in order to encourage increased levels of engagement and, potentially, reduce the risk of some of their followers from churning, thus hopefully increasing their social media following over time.

10.4 Potential Limitations and Implications for Future Research

The previous chapter of this thesis included discussion of four areas of potential limitations of the studies presented in prior chapters, and how this may inform future work in this area. This future work can be broadly categorised into two areas of research.

The first area is to better understand the numerous and complex motivational factors for individual users to begin engaging with a given social media account. In the studies presented in this thesis, engagement with particular accounts was analysed, and many users that follow these accounts did not engage at all within the collected datasets. It could be that these users saw no need to actively engage with the given account, or that they do not actively engage with any (or very few) accounts on the platform. If users are selective in which accounts they actively engage with, understanding this behaviour may help in understanding what may motivate them to begin engaging with an organisation's account. If, on the other hand, users make use of social media in an entirely passive way (yet still use the platform), then understanding how they can be best served through social media may benefit the organisation in their 'offline' performance, if not in their online performance.

The second area is to better understand the factors surrounding users and their motivations for remaining in the audience of a given social media account. As demonstrated in the studies presented here, both a user's positioning in the network and (to a much lesser extent) their active engagement with content can be used to model their potential to 'churn'. However, there is still much room for improvement in these models, indicating that there are more contributing factors. A greater understanding of what these factors are will allow organisations to further develop and refine their online presence to better meet the needs of their existing audiences, and by extension encourage the growth of the audience – ultimately to the benefit of the organisation.

10.5 Summary

This chapter concluded the thesis by reviewing the aims and objectives outlined at the beginning, noting how each was met within the various studies, and in doing so, how they informed later elements of the research, and the implications for future research. Taken as a whole, the research presented here has contributed to a greater understanding of how organisations make use of social media, and the audience response to this use of social media. Through modelling and understanding user engagement with content, organisations can tailor future social media content and campaigns to encourage continued and greater engagement from their audience. Further, potential 'churn' in an audience can be identified through the modelling of the audience network, as well as the audience's prior engagement behaviours. Such information can be used to further tailor social media activities, in an effort to retain individuals deemed to be of value. Implications for future work have also been identified – including the importance of continued work in this area, and a growing need for organisations to better understand their own audiences.

Appendix A

Ethics Applications

This appendix includes all available documentation relating to the ethical approval processes undertaken during this research. Documents pertaining to approval at the University of Lincoln are first included, followed by documentation from Northumbria University.

EA1

[doc version 09.02]

Ethical Approval Form: Library/Desk/Lab/Studio-based Research Projects

This form must be completed for each piece of research activity whether conducted by academic staff, research staff, graduate students or undergraduates. Applications by students must be endorsed by an academic member of staff acting as Principal Investigator/supervisor. The completed form must be sent to the designated Ethics Committee within the College.

Please complete all sections. If a section is not applicable, write N/A.

1 Name of Applicant	Jamie Mahoney
2 School	School of Computer Science
3 Position in the University	MPhil / PhD Student
4 Role in relation to this research	Researcher
5 Name(s) of collaborators/co-workers and their relationship to the project (e.g. supervisor, assistant etc.)	*Name, and role in project:* 1. Prof. Shaun Lawson - Supervisor
6 Brief statement of main Research Question or Project Title	Social Media Analytics in the Retail Sector
7 Ethical checklist	Does the research involve living human participants, or human tissue? **Yes** ☒ **No** ☐ *If you answered "yes", submit form EA2 for Ethical Approval.*
	Does the research involve living animals, or animal tissue? **Yes** ☐ **No** ☒ *If you answered "yes", submit form EA3 for Ethical Approval.*
	Does the research involve confidential data, or data not in the public domain? **Yes** ☐ **No** ☒
	Does the project potentially put you or your collaborators at physical or psychological risk? **Yes** ☐ **No** ☒
	Could the topic or results of this research be seen as illegal, or attract legal action against the University from an outside agency? **Yes** ☐ **No** ☒
	Could the topic or results of this research attract unwelcome media attention, or affect the reputation or standing of the University? **Yes** ☐ **No** ☒
	Could the topic, results or conduct of this research be regarded as offensive, immoral or destructive by some reasonable people? **Yes** ☐ **No** ☒
	Does this research need to be undertaken under a relevant professional code of conduct? **Yes** ☐ **No** ☒
	Are there any potential conflicts of interest in conducting this research, including financial gain for the researchers, or for individuals or external organizations affiliated with the researchers? **Yes** ☐ **No** ☒
	Are there any factors inhibiting the application of the University's ethical guidelines, including those on proper treatment of data, research design and publication of results? **Yes** ☐ **No** ☒
	Does the research require the approval of any external body? **Yes** ☐ **No** ☒

	If the answer to all questions above is "No", you may complete section 8 to certify that there are no ethical issues, submit this form to the relevant Ethics Committee, and proceed with the research immediately. You accept professional responsibility for this decision, and if unsure should instead submit to the Committee. ***If the answer to any of the above questions is "Yes", complete the rest of the form, submit to the relevant Ethics Committee, and await approval before proceeding with the research. Answering "Yes" does not necessarily imply that the research is problematic, only the Ethics Committee needs to consider the research to ensure that it can proceed, and that the research design conforms to best practice.***
8 Self certification of Ethical Review	*Having reviewed the ethical implications of this research, I certify that there are no issues requiring Ethical Approval. I certify that the research will be carried out in compliance with the University's ethical guidelines for library/desk/laboratory/studio-based research, with Health and Safety regulations, and with all other relevant University policies and procedures. If there are any changes to the research requiring ethical clearance, I shall apply for such clearance before continuing with the research.* Signed: Principal Investigator ***Note.*** *This section must be endorsed by the member of academic staff responsible for the project. In the case of research by students, the supervising member of academic staff must sign. The signed form should then be submitted to the relevant Ethics Committee within the College, and the research may proceed.*
9 Does the research comply with the University's key ethical principles for library/desk/lab/studio-based research ?	**Yes** ☒ **No** ☐ If "No", provide an ethical justification for your project and explain why you wish to continue with the research in breach of normal ethical principles:-
10 If applicable, please state the relevant professional code(s) under which the research is being conducted and confirm compliance	
11 Does this research require the approval of an external body ?	**Yes** ☐ **No** ☒ **If "Yes", please state which body:-**
12 Has ethical approval already been obtained from that body ?	**Yes** ☐ **-Please append documentary evidence to this form.** **No** ☐ **If "No", please state why not:-** **Please note that any such approvals must be obtained and documented before the project begins.**

13 If there are any other ethical issues, to which the attention of the approving committee should be drawn, please state them in this section, and explain how you have taken the issues into account, so that the research should be approved. Please consult the University's ethical guidelines for advice. **Please also include here, or attach separately, a brief description of the research, to allow the approving committee to reach judgement.**	See associated submitted EA2

<u>APPLICANT SIGNATURE</u>

I hereby request ethical approval for the research as described above.
I certify that I have read the University's ethical guidelines for library/desk/laboratory/studio-based research.

____________________________________ ________________
Applicant Signature **Date**

____
PRINT NAME

FOR STUDENT APPLICATIONS ONLY –
Academic Support for Ethics
Academic support should be sought prior to submitting this form to the designated Ethics Committee within the Faculty.

.

• ***Undergraduate / Postgraduate Taught application***	***Academic Member of staff nominated by the School (consult your project tutor)***
• ***Postgraduate Research Application***	***Director of Studies***

I support the application for ethical approval

______________________________ ______________
Academic / Director of Studies Signature Date

____
PRINT NAME

FOR COMPLETION BY THE DESIGNATED ETHICS COMMITTEE WITHIN THE COLLEGE

Please select ONE of A, B, C or D below:

☐ **A. Ethical approval to this research.**

☐ **B. Conditional ethical approval to this research.**

10 Please state the condition (inc. date by which condition must be satisfied if applicable)	

☐ **C. Ethical approval cannot be given to this research but the application is referred on to the University Research Ethics Committee for higher level consideration.**

11 Please state the reason	

☐ **D. Ethical approval cannot be given to this research and it is recommended that the research should not proceed.**

12 Please state the reason, bearing in mind the University's ethical framework, including the primary concern for Academic Freedom.	

Signature of the Chair of the designated ethics committee within the College

Signature: ________________________________Date:________________________________

Chair of________________________________
Key ethical guidelines for library/desk/laboratory/studio-based research

The University of Lincoln has drawn up the following key principles for researchers engaged in library/desk/laboratory/studio-based projects in order to promote high professional standards. They should be read alongside the University's Ethical Principles for Conducting Research with Humans and Other Animals, and operate as part of the University's Ethical Framework.

- *Non-falsification of data:* Researchers have an ethical obligation to refrain from tampering with data. Thus questionnaire responses, experimental observations and data analyses should not be fabricated, altered nor discarded. In addition, researchers have a responsibility to exercise reasonable care in processing data to ensure no errors affect the results.
- *Ethics of reporting research:* Researchers are obliged to give full and proper attribution of ideas: presenting the words, data or ideas of another person as your own without properly citing them amounts to plagiarism. This is not only misconduct but can also be an infringement of copyright, amounting to theft of intellectual property.
- *Ethics and research design:* Researchers should be open to a range of methods: failure to consider and evaluate alternative methods and tools for the collection of data may be regarded as too overtly biased. All appropriate steps should be taken to ensure that no samples are obtained from unethical sources e.g. illegal databases; unregistered suppliers of samples from humans or other animals.
- *Authorship credit:* Only those researchers who are significant contributors to a research project should be given authorship credit. A "significant contributor" might be described as a person playing a major role in conceptualising, analysing or writing the final document. Ideally, all those involved in the research project should decide upon the order of authorship. Usually, the first author is the one who has made the biggest contribution.
- *Conflict of interest:* Researchers should be aware of the potential influence of personal or commercial interests on their work and take all practical measures to ensure that information is presented without distortion.
- *The principle of beneficence:* Researchers are required to protect individuals by seeking to maximise anticipated benefits and minimise possible harms. It is therefore necessary to examine carefully the design of the study and its risks and benefits including, in some cases, identifying alternative ways of obtaining the benefits sought from the research. Research risks must always be justified by the expected benefits of research.
- *Professional codes:* Researchers should undertake research legally and in accordance with any relevant professional codes of conduct.
- *Personal information:* Researchers should anonymise information which relates to individuals when they have not obtained informed consent, unless there is a clear justification to the contrary. They should also be aware of the impact of wider public dissemination of their work and the impact this might have on any individual or group of individuals. If it is anticipated that it might cause distress, it is essential to demonstrate that the benefits outweigh this risk.

EA2

Ethical Approval Form: Human Research Projects

Please word-process this form, handwritten applications will not be accepted

UNIVERSITY OF LINCOLN

This form must be completed for each piece of research activity whether conducted by academic staff, research staff, graduate students or undergraduates. The completed form must be approved by the designated authority within the College.

Please complete all sections. If a section is not applicable, write N/A.

1 Name of Applicant	Jamie Mahoney	
	School: School of Computer Science	College: College of Science
2 Position in the University	MPhil / PhD Student	
3 Role in relation to this research	Researcher	
4 Brief statement of main Research Question	Social Media Analytics in the Retail Sector	
5 Brief Description of Project	Social media presents new digital interaction opportunities and challenges to urban retail locations such as shopping malls, centres and streets. Platforms such as Facebook and Twitter facilitate online communication with, and between, customers that is not possible through traditional media and marketing techniques. This project will utilise data from social media websites (such as Twitter) to understand user behaviour when using these websites, and the communities that are formed during this use, with specific reference to the retail sector.	
	Approximate Start Date: Jan 2014	Approximate End Date: Jan 2018
6 Name of Principal Investigator or Supervisor	Prof. Shaun Lawson (Supervisor)	
	Email address: [redacted]	Telephone: [redacted]
7 Names of other researchers or student investigators involved	Jamie Mahoney – Sole Researcher – Doctoral Study	
8 Location(s) at which project is to be carried out	Research conducted in Lincoln on Brayford Campus	

9 Statement of the ethical issues involved and how they are to be addressed –including a risk assessment of the project based on the vulnerability of participants, the extent to which it is likely to be harmful and whether there will be significant discomfort. **(This will normally cover such issues as whether the risks/adverse effects associated with the project have been dealt with and whether the benefits of research outweigh the risks)**	This project will utilise data from social media websites (such as Facebook and Twitter) to understand user behaviour when using these websites, and the communities that are formed as a part of this usage. The research is desk-based, but involves collecting and analysing data generated by users of social media websites. The legitimacy of using, for research purposes, publicly available postings, media and other content from social media sites is an open question in the social computing research community (Zimmer, 2010). The current consensus, however, is that such an approach is acceptable, so long as researchers take adequate steps to protect the anonymity of those whose data is included in the data sample, in any subsequent publications. The British Psychological Society's recently issued 'Ethics Guidelines for Internet-Mediated Research' provides guidance for situations in which it is not necessary to seek out individuals to gain informed consent from them. As the vast majority of data posted to Twitter should be regarded as being in the public domain (Twitter, 2013), it is reasonable to argue that there is no likely perception or expectation of privacy and as such, use of this data without gaining consent can be justified (British Psychological Society, 2013). The exceptions to this are, of course, direct messages between users (which are often seen as being private) and posts made from accounts that have access restricted to them (private accounts). This data is not accessible through the data collection mechanisms made available by Twitter, and as such the data will not be collected during the course of this project. In order to reduce the risk of social media users being identified in any subsequent publications, two main procedures will be followed. Firstly, the account names of people included in the datasets will not be published; where necessary, pseudonyms or other identification methods will be used. This will avoid an identified potential risk of having people exposed to judgement or ridicule for opinions expressed via social media (NatCen Social Research, 2014). Secondly, examples of posts will be paraphrased or combined with others, where necessary, to avoid situations where search engines and other tools could be used to identify individual accounts from the text of posts that they have shared (British Psychological Society, 2013). <u>References</u> British Psychological Society (2013). *Ethics Guidelines for Internet-Mediated Research.* INF206/1.2013. Leicester: Author. Available from: http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf NatCen Social Research (2014) *Research Using Social Media; Users' Views*. Available from: http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf Twitter (2013) *Twitter Privacy Policy.* Available from: https://twitter.com/privacy Zimmer, M. (2010) "But the data is already public": on the ethics of research in Facebook. Ethics and Information Technology 12, 4 (2010), 313-325.

Ethical Approval From Other Bodies

10 Does this research require the approval of an external body ?	**Yes** ☐	**No** ☒
	If "Yes", please state which body:-	
11 Has ethical approval already been obtained from that body ?	**Yes** ☐ **-Please append documentary evidence to this form.** **No** ☐ **If "No", please state why not:-** **Please note that any such approvals must be obtained and documented before the project begins.**	

APPLICANT SIGNATURE

I hereby request ethical approval for the research as described above.
I certify that I have read the University's ETHICAL PRINCIPLES FOR CONDUCTING RESEARCH WITH HUMANS AND OTHER ANIMALS.

____________________________________ ________________

Applicant Signature **Date**

____

PRINT NAME

FOR STUDENT APPLICATIONS ONLY –
Academic Support for Ethics

Academic support should be sought prior to submitting this form to the designated Ethics Committee within the Faculty .

Undergraduate / Postgraduate Taught application	***Academic Member of staff nominated by the School (consult your project tutor)***
Postgraduate Research Application	***Director of Studies***

I support the application for ethical approval

____________________________________ ________________

Academic / Director of Studies Signature   Date

____
PRINT NAME

<u>FOR COMPLETION BY THE DESIGNATED ETHICS COMMITTEE WITHIN THE COLLEGE</u>

Please select ONE of A, B, C or D below:

☐ **A. Ethical approval is given to this research.**

☐ **B. <u>Conditional</u> ethical approval is given to this research.**

Please state the condition (inc. date by which condition must be satisfied if applicable)	

☐ **C. Ethical approval cannot be given to this research but the application is referred on to the University Research Ethics Committee for higher level consideration.**

Please state the reason	

☐ **D. Ethical approval cannot be given to this research and it is recommended that the research should <u>not</u> proceed.**

Please state the reason, bearing University's ethical framework, ıe primary concern for Academic	

Signature of the Chair of the designated Ethics Committee within the College

____________________________________   ______________
Signature   **Date**

Chair of____

Faculty of Engineering and Environment

RESEARCH PROJECT: ETHICS REGISTRATION AND APPROVAL FORM

Section One: Registration *[To be completed by researcher]*

Title of research project/dissertation	*Social Media Analytics in the Retail Sector*

Researcher's name	Jamie Mahoney
Student number (if applicable)	[redacted]

Please only complete the following if researcher is a student:

Programme of study	PhD

Academic Year	16/17

Module code (if applicable)	

Principal Supervisor or Module Tutor	Prof Shaun Lawson

Start Date	Jan 2017

Brief outline of research topic:
Social media presents new digital interaction opportunities and challenges to urban retail locations such as shopping malls, centres and streets. Platforms such as Facebook and Twitter facilitate online communication with, and between, customers that is not possible through traditional media and marketing techniques.

This project will utilise data from social media websites (such as Twitter) to understand user behaviour when using these websites, and the communities that are formed during this use, with specific reference to the retail sector

Short description of proposed research methods including identification of participants:
A number of retail locations will be identified, each of which will have an active presence on online social media. Each of these accounts will have public 'followers' – i.e. individuals that demonstrate an interest in these accounts. Secondary data for each of these will be collected, mainly their public engagement with each of the retail locations, along with public social graph data which demonstrates the online relationships between these accounts. Through qualitative analysis, engagement behaviours will be profiled, and the impact of this engagement on the social graph can be determined.

Ethical considerations in the research project	YES	NO
1. Does your research involve an external organisation or partner?	☐	☒
2. Does your research involve human participants?	☐	☒
3. If yes to Q.2, will you inform the participants about the research?	☐	☒
4. Will you obtain their consent using the standard consent form?	☐	☒
5. Is any deception involved?	☐	☒
6. Do any participants constitute a 'vulnerable group'? (refer to definition of Vulnerable People)	☐	☒
7. Will the research involve the following information?		
Commercially sensitive	☐	☒
Personally sensitive	☐	☒
Politically sensitive	☐	☒
Legally sensitive	☐	☒
8. Is the research likely to have any significant environmental impacts?	☐	☒
9. Are there likely to be any risks for the participants in your research?	☐	☒
10. Are there likely to be any risks for you in conducting the research?	☐	☒
11. If yes [to 5, 6, 7, 8, 9 or 10 above] have you identified steps to address the issues and mitigate any risks to participants, yourself or the environment?	☐	☐

Statement to explain how any issues identified above will be addressed and what steps will be taken to mitigate such risks or adverse impacts

This project will utilise data from social media websites (such as Facebook and Twitter) to understand user behaviour when using these websites, and the communities that are formed as a part of this usage. The research is desk-based, but involves collecting and analysing data generated by users of social media websites.

The legitimacy of using, for research purposes, publicly available postings, media and other content from social media sites is an open question in the social computing research community (Zimmer, 2010).

The current consensus, however, is that such an approach is acceptable, so long as researchers take adequate steps to protect the anonymity of those whose data is included in the data sample, in any subsequent publications.

The British Psychological Society's recently issued 'Ethics Guidelines for Internet-Mediated Research' provides guidance for situations in which it is not necessary to seek out individuals to gain informed consent from them. As the vast majority of data posted to Twitter should be regarded as being in the public domain (Twitter, 2013), it is reasonable to argue that there is no likely perception or expectation of privacy and as such, use of this data without gaining consent can be justified (British Psychological Society, 2013). The exceptions to this are, of course, direct messages between users (which are often seen as being private) and posts made from accounts that have access restricted to them (private accounts). This data is not accessible through the data collection mechanisms made available by Twitter, and as such the data will not be collected during the course of this project.

In order to reduce the risk of social media users being identified in any subsequent publications, two main procedures will be followed. Firstly, the account names of people included in the datasets will not be published; where necessary, pseudonyms or other identification methods will be used. This will avoid an identified potential risk of having people exposed to judgement or ridicule for opinions expressed via social media (NatCen Social Research, 2014). Secondly, examples of posts will be paraphrased or combined with others, where necessary, to avoid situations where search engines and other tools could be used to identify individual accounts from the text of posts that they have shared (British Psychological Society, 2013).

References

British Psychological Society (2013). Ethics Guidelines for Internet-Mediated Research. INF206/1.2013. Leicester: Author. Available from: http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf

NatCen Social Research (2014) Research Using Social Media; Users' Views. Available from: http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf

Twitter (2013) Twitter Privacy Policy. Available from: https://twitter.com/privacy

Zimmer, M. (2010) "But the data is already public": on the ethics of research in Facebook. Ethics and Information Technology 12, 4 (2010), 313-325.

Ethical category of research project

Based on the above Ethical Considerations and with reference to the University's Ethical Scrutiny Risk Assessment tool identify the Ethical category of your research project (refer to http://www.northumbria.ac.uk/static/5007/respdf/riskassesmenttool for further guidance):

[Please tick as appropriate]

Red	☐	vulnerable participants; human tissue; sensitive data; risks to participants & researchers etc.
Amber	☐	human participants requiring informed consent; commercially sensitive information etc.
Green	☒	no participants involved; secondary data only; no sensitive data

I have read the University and the Faculty Ethics Policy and Procedures and confirm that the answers I have given above are correct. Where issues arise under items 5, 6, 7, 8, 9 or 10 [above] I have described in writing how I intend to approach these issues in the research.

Researcher's signature **J Mahoney**................................

Date **17/01/2017**................................

Section 1 Ethics Registration to be submitted to Principal Supervisor or Module Tutor

Section Two: Approval

Supervisor/Module Tutor's name	

Ethical approval *[Please tick as appropriate]*

Green - Ethical approval is given without conditions	☒
Amber - Ethical approval is given with the following conditions: Information to be provided to all participants Participant consent to be obtained using the standard Research Participant Consent Form or otherwise in accordance with Faculty procedures Data to be stored and destroyed securely in accordance with University guidelines Adherence to DPA Anonymity to be provided to participants Commercial confidentiality to be provided to organisations(s) Other (please state):	 ☐ ☐ ☐ ☐ ☐ ☐ ☐
Red - Project is referred to FREC for approval	☐

Name & role of approving member of staff 1: ……………………………………….

Signature ……………………………………….

Date ……………………………………….

Outcome of FREC referral – Decision, minute and date of meeting; identify two signatories

Research data

All of the questions on this form must be completed and the form included as part of your research ethics submission.

To complete the form, please type your responses into the text boxes, which will expand to accommodate the information you provide. For questions with the option of 'yes/no', please click on the appropriate box and type '**X**'.

Depending on your research study, you may also need to include supporting documentary evidence as part of this form. Please refer to the research ethics guidelines for your academic school for information about the type of evidence you need to provide.

Research ethics number: *From Research Information tab of Online Form*	RE-EE-16-170116-587cbbafdea7c
Project title:	Social Media Analytics in the Retail Sector

1. Data sources

Provide details of the data sources you will use during this research study.

Public data made available through online social media, namely Twitter and Instagram

Will this research study use data that is commercially sensitive? ☐ Yes ☒ No

If yes: State clearly where you are sourcing your data from and provide the job title and contact details (address and telephone number) of the person you will approach for permission to use the data in the study. If you have already received permission to use the data, please include a copy of the letter or email confirming access to the data under 'Supporting Documentary Evidence' in the File Manager tab of the online form.

Will this research study use data that is covered by the UK Official Secrets Act?

☐ Yes ☒ No

If yes: State clearly where you are sourcing your data from and provide the job title and contact details (address and telephone number) of the person you will approach for permission to use the data in the study. If you have already received permission to use the data, please include a copy of the letter or email confirming access to the data under 'Supporting Documentary Evidence' in the File Manager tab of the online form.

2. Research environment

Provide details of the different locations or venues where you will be undertaking the research study. This may include: the University library; your office on campus; or off-campus locations such as your home or the organisation or institution where the data sources for the study are stored.

Office on Campus, Home

Provide details of any risks that the study may pose to researcher safety and the measures you will take to minimise these risks. Depending on the nature of the research study this may include: lone working; working with machinery; the control of substances hazardous to health (COSHH).

Potential for lone working – university and faculty lone working procedures will be followed as and when necessary

3. Data storage

Describe the arrangements for the secure transport and storage of data collected during the study and the measures that will be taken to back-up data and information.

Data will be collected using tools made publicly available by the various social media platforms. Any collected data will be public data, and will be stored securely on hard drives, including any backups that are made.

4. Data Retention and Disposal

Describe the arrangements for the secure retention and/or disposal of data when the research study is complete.

Once the study is complete, datasets will be completely anonymised, replacing user IDs (for example), with randomly assigned numbers.

Supporting Documentary Evidence

Depending on your research study, you may need to include supporting documentary evidence with this form. The documentary evidence you need to provide depends on:

- Your research study

- The research ethics guidelines of your academic school

Do you have supporting documentary evidence? ☐ Yes ☒ No

If yes: Your supporting documentary evidence should be upload under 'Supporting Documentary Evidence' in the File Manager tab of the online form.

Appendix B

Description of Retail Locations

The following retail locations have been used as case studies in the studies presented within this thesis. Details of each location, their web presence, and social media activity are provided.

B.1 Bluewater Shopping Centre, Kent

Bluewater Shopping Centre is an out-of-town shopping centre in Kent (see Figure B.1), based just outside of the M25 orbital motorway which surrounds London. Having first opened in 1999, Bluewater now hosts over 300 retail stores and services, with a total retail floor area of over 155,000m^2, making it currently the fifth largest shopping centre in the United Kingdom.

Maintaining an active web presence (Figure B.2), Bluewater also actively promotes its social media presence on Twitter, YouTube, Facebook, and Instagram; example social media posts are shown in Figures B.3 and B.4.

B.2 Boxpark Shoreditch, London

Boxpark locations, including Shoreditch, are retail parks constructed from refitted shipping containers (see Figure B.5). Originally opened in 2011, Boxpark Shoreditch was originally intended to operate for only five years, but is still operating, as of late 2019, featuring 19 restaurants and 27 shops. The

Figure B.1: Bluewater Shopping Centre, Kent. Source: https://bluewater.co.uk/

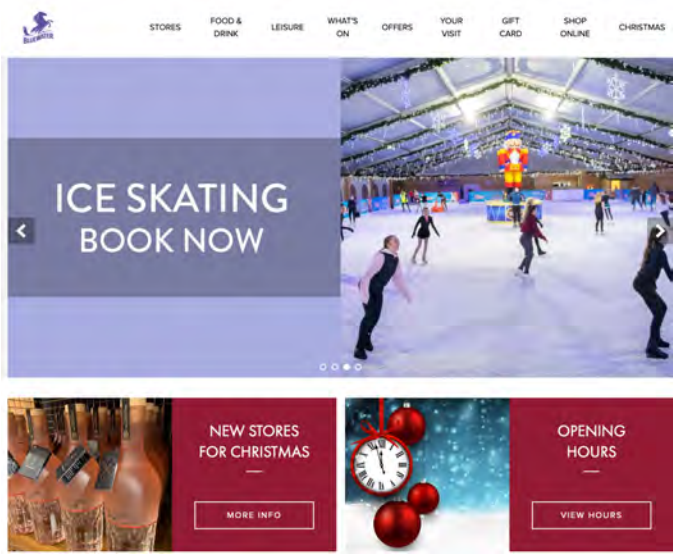


Figure B.2: Bluewater Shopping Centre's web presence. Source: https://bluewater.co.uk/

Figure B.3: An example Facebook post from Bluewater Shopping Centre. Source: https://www.facebook.com/Bluewater/

Figure B.4: An example Instagram post from Bluewater Shopping Centre. Source: https://www.instagram.com/bluewatershopping/

Boxpark concept now includes retail locations in Croydon and Wembley. Other Boxpark locations are planned for other UK cities, and other countries.

Boxpark maintains an active web presence, including details for each of its locations (see Figure B.6), and maintains a social media presence (with individual accounts for each location) on Twitter, Facebook, YouTube and Instagram. Example social media posts are also shown in Figures B.7 and B.8.

B.3 Brent Cross Shopping Centre, London

Brent Cross (shown in Figure B.9) was the UK's first out-of-town style shopping centre, first opening in 1976. Covering a total retail floor area of over 74,000m^2, Brent Cross has also been used as a filming location in multiple films, including *Tomorrow Never Dies*, the 1997 James Bond film.

Brent Cross maintains an active web presence (see Figure B.10, and has a social media presence on a range of platforms - Twitter, Facebook, and Instagram. Example posts can be seen in Figures B.11 and B.12.

B.4 Bullring & Grand Central Shopping Centre, Birmingham

When combined (they are physically connected via a footbridge), Bullring & Grand Central (shown in Figure B.13) becomes the UK's largest city centre shopping centre, housing approximately 200 retailers and services - including Selfridges and one of the largest Debenhams in the UK. As with all of

Figure B.5: Boxpark Shoreditch, London. Source: https://www.boxpark.co.uk

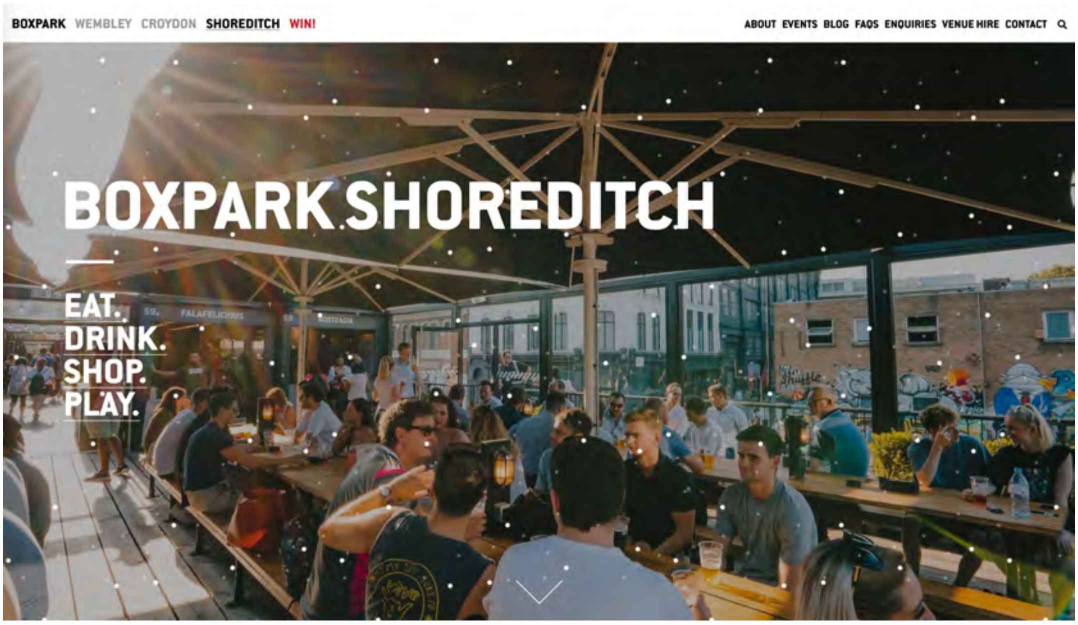


Figure B.6: Bluewater Shopping Centre's web presence. Source: https://www.boxpark.co.uk

Figure B.7: An example Facebook post from Boxpark Shoreditch. Source: https://www.facebook.com/boxparkshoreditch.

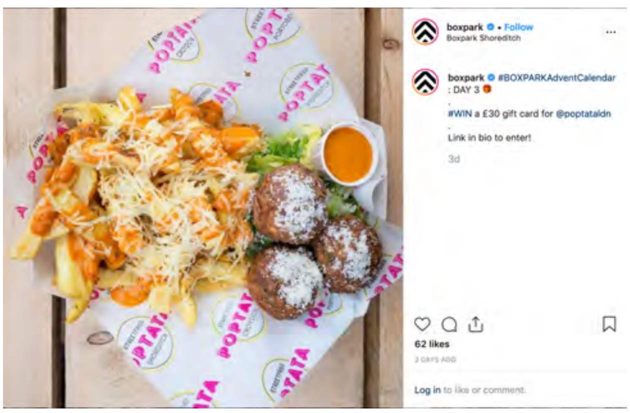


Figure B.8: An example Instagram post from Boxpark Shoreditch. Source: https://www.instagram.com/boxpark/.

Figure B.9: Brent Cross, London. Source: https://en.wikipedia.org/wiki/Brent˙Cross˙Cricklewood

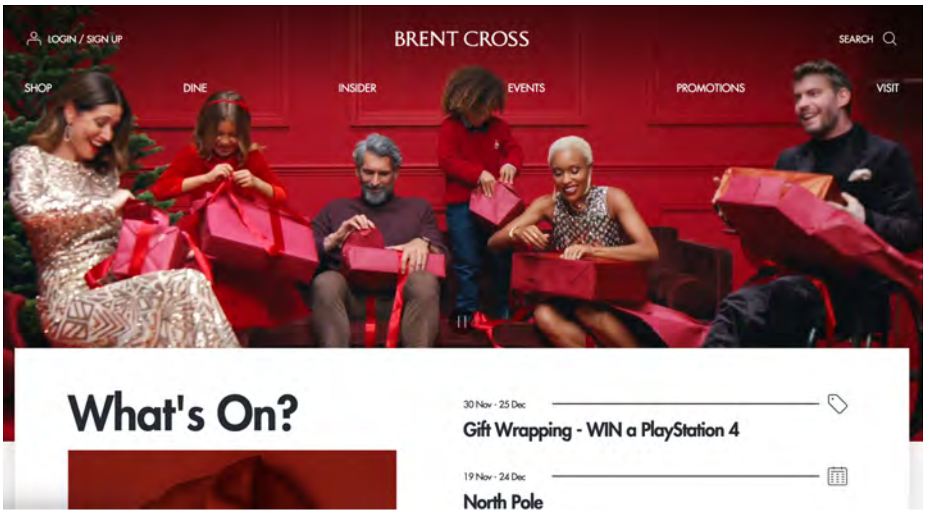


Figure B.10: Brent Cross's web presence. Source: https://www.brentcross.co.uk

Figure B.11: An example Twitter post from Brent Cross. Source: https://www.twitter.com/brentcross_sc.

Figure B.12: An example Instagram post from Brent Cross. Source: https://www.instagram.com/brentcross_sc.

Figure B.13: Bullring & Grand Central, Birmingham. Source: http://www.bullring.co.uk

the other retail locations detailed here, Bullring & Grand Central maintains an active online presence, both in terms of a website (see Figure B.14) and a social media presence on Twitter, Facebook, and Instagram. Example posts from these can be seen in Figures B.15 and B.16.

B.5 Carnaby London

Carnaby Street, a pedestrianised shopping street in Central London (shown in Figure B.17), developed a reputation in the 1960's (along with the surrounding streets), attracting independent fashion boutiques, designers, and music venues. Now presented as a collection of streets, the Carnaby area offers over 100 retailers, and a further 60 restaurants, bars and cafes. Carnaby maintains an active web presence (see Figure B.18), along with a presence on a wide range of social media platforms - Facebook, Twitter, Instagram, YouTube and Snapchat. Some example posts from these platforms can be seen in Figures B.19 and B.20

B.6 Covent Garden, London

Covent Garden (shown in Figure B.21) is an area in London that offers a range of retailers, restaurants, and entertainment venues. While many other retail locations included within this thesis have very clear geographical constraints (such as being within a self-contained building), Covent Garden is less

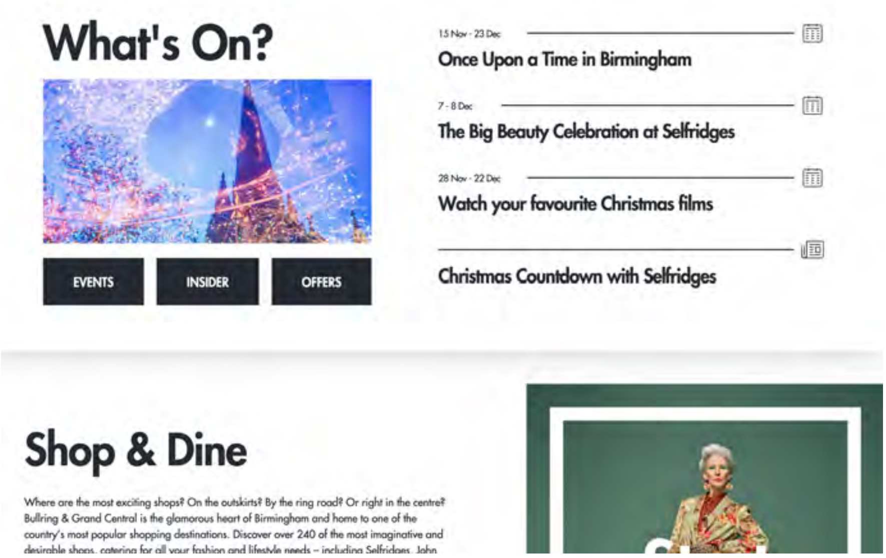


Figure B.14: Bullring & Grand Central's web presence. Source: https://www.bullring.co.uk

Figure B.15: An example Twitter post from Bullring & Grand Central. Source: https://www.twitter.com/bullring.

Figure B.16: An example Instagram post from Bullring & Grand Central. Source: https://www.instagram.com/bullring.

Figure B.17: Carnaby Street, London. Source: http://www.carnaby.co.uk

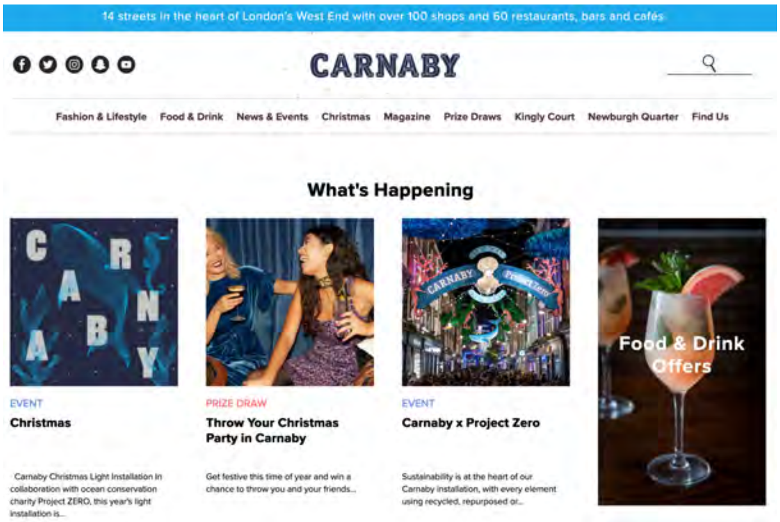


Figure B.18: Carnaby London's web presence. Source: https://www.carnaby.co.uk

Figure B.19: An example Facebook post from Carnaby London. Source: https://www.facebook.com/carnabylondon.

Figure B.20: An example Instagram post from Carnaby London. Source: https://www.instagram.com/carnabylondon.

Figure B.21: Covent Garden, London. Source: http://www.coventgarden.london

so. In its entirety, the Covent Garden Estate covers approximately 50,000m^2. Alongside a website (see Figure B.22), a social media presence is also maintained on Facebook, Twitter, Instagram, YouTube, and Pinterest. A selection of example social media posts are shown in Figures B.23 and B.24.

B.7 intu Metrocentre, Gateshead

Currently branded as intu Metrocentre, the Metrocentre (shown in Figure B.25) and surrounding retail parks offer approximately 500 retailers and associated services, covering a total retail floor space of over 192,000m^2, making it the second-largest shopping centre in the United Kingdom. Originally opening in 1986, the Metrocentre now houses six 'anchor tenants' - including Debenhams and Marks & Spencer. Offering over 8,000 parking spaces, and being well-served by public transport, intu reports that the Metrocentre has an annual footfall of over 20 million visitors [99].

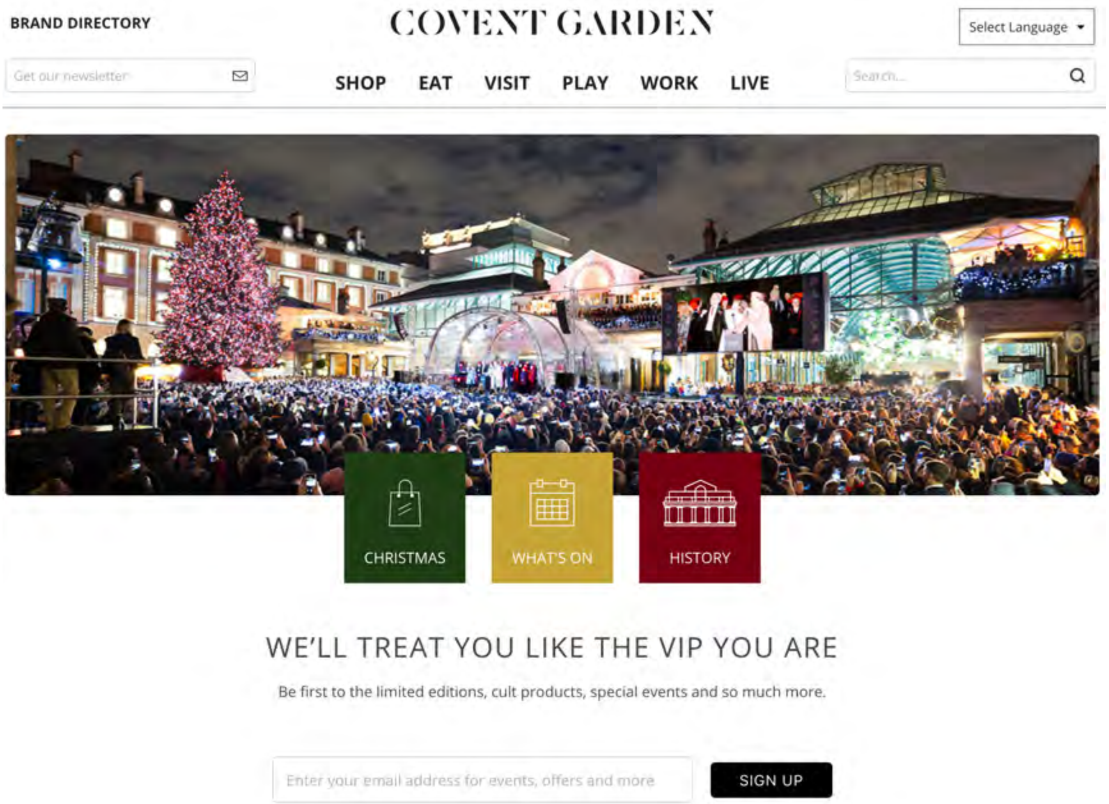


Figure B.22: Covent Garden's web presence. Source: https://www.coventgarden.london

Figure B.23: An example YouTube post from Covent Garden. Source: https://www.youtube.com/channel/UCnWdDyr3ojs_9X7YzFJPNsw.

Figure B.24: An example Instagram post from Covent Garden. Source: https://www.instagram.com/coventgardenldn.

Figure B.25: intu Metrocentre, Gateshead. Source: http://www.facebook.com/intuMetrocentre

intu Metrocentre currently maintains an extensive web presence (see Figure B.26), as well as an active presence on multiple social media platforms - including Facebook, Instagram, Pinterest, and Twitter; example social media posts from these accounts are shown in Figures B.27 and B.28.

B.8 London West End

'London West End', including Leicester Square (shown in Figure B.29), is well known for its theatres, restaurants and retailers, and attracts millions of visitors per year. Branded as either London West End, or Leicester Square (as the main focal point for the West End), an active web (see Figure B.30) and social media presence is maintained - across Facebook, Twitter, Instagram and YouTube. Examples of the location's web presence are shown in Figures B.31 and B.32.

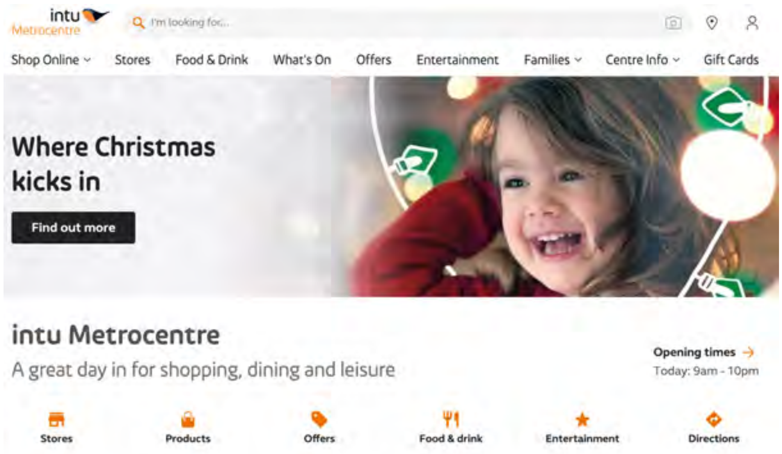


Figure B.26: intu Metrocentre's web presence. Source: https://www.intu.co.uk/metrocentre

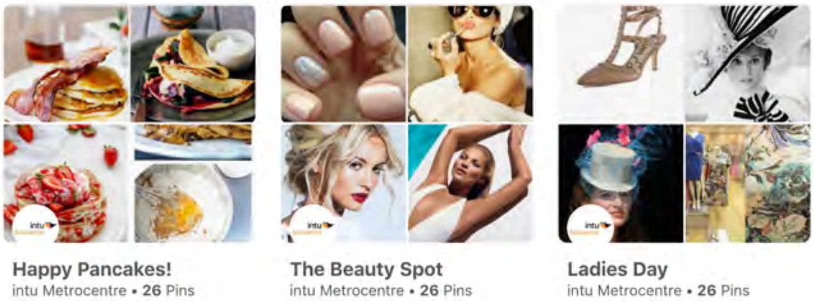


Figure B.27: An example of Pinterest posts from intu Metrocentre. Source: https://www.pinterest.com/metrocentre

.

Figure B.28: An example Facebook post from intu Metrocentre. Source: https://www.facebook.com/intuMetrocentre

.

Figure B.29: London West End / Leicester Square. Source: http://www.facebook.com/discoverlsq

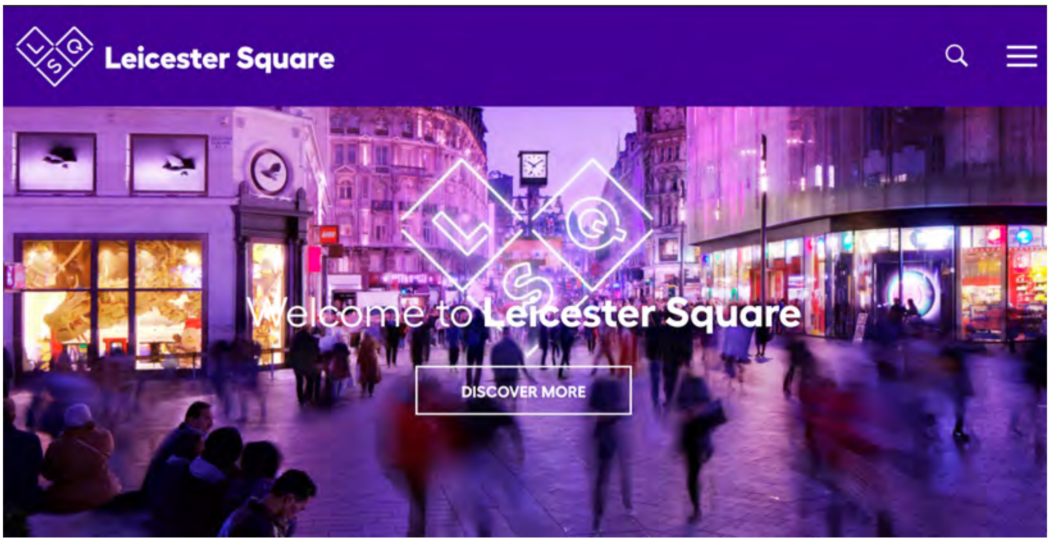


Figure B.30: London West End / Leicester Square's web presence. Source: https://www.leicestersquare.london

Figure B.31: An example of London West End / Leicester Square's Twitter presence. Source: https://twitter.com/londonwestend.

Figure B.32: An example Facebook post from London West End / Leicester Square. Source: https://www.facebook.com/discoverlsq

.

Figure B.33: Oxford Street, London. Source: http://www.facebook.com/OxfordStW1

B.9 Oxford Street, London

Oxford Street (shown in Figure B.33), Europe's busiest shopping street, attracts around 500,000 visitors daily. Approximately 1.2 miles long, the street is home is to around 300 different retailers, each attracting their own audiences. Oxford Street is home to the 'flagship' stores of several retailers - including Selfridges, HMV and Debenhams.

Alongisde their main website (see Figure B.34), Oxford Street maintains an active presence on Twitter, Facebook, and Instagram. Example posts from their social media accounts are shown in Figures B.35 and B.36.

B.10 Regent Street, London

Just under one mile long, Regent Street crosses both Oxford Street and Piccadilly Circus in the West End of London (see Figure B.37). First developed in the early 19th century, the street is now known as the home of many retailers, including the 'flagship' stores of Liberty, Hamleys, and Apple. As well as being the home of several major retailers, Regent Street also hosts the Regent Street Festival each year, when the street is completely closed to vehicular traffic. For example, in the summer of 2004, an estimated 500,000 people visited Regent Street and the surrounding streets to witness a parade of Formula One cars.

Figure B.34: Oxford Steet, London's web presence. Source: https://www.oxfordstreet.co.uk

Figure B.35: An example Twitter post from Oxford Street, London. Source: https://twitter.com/OxfordStreetW1.

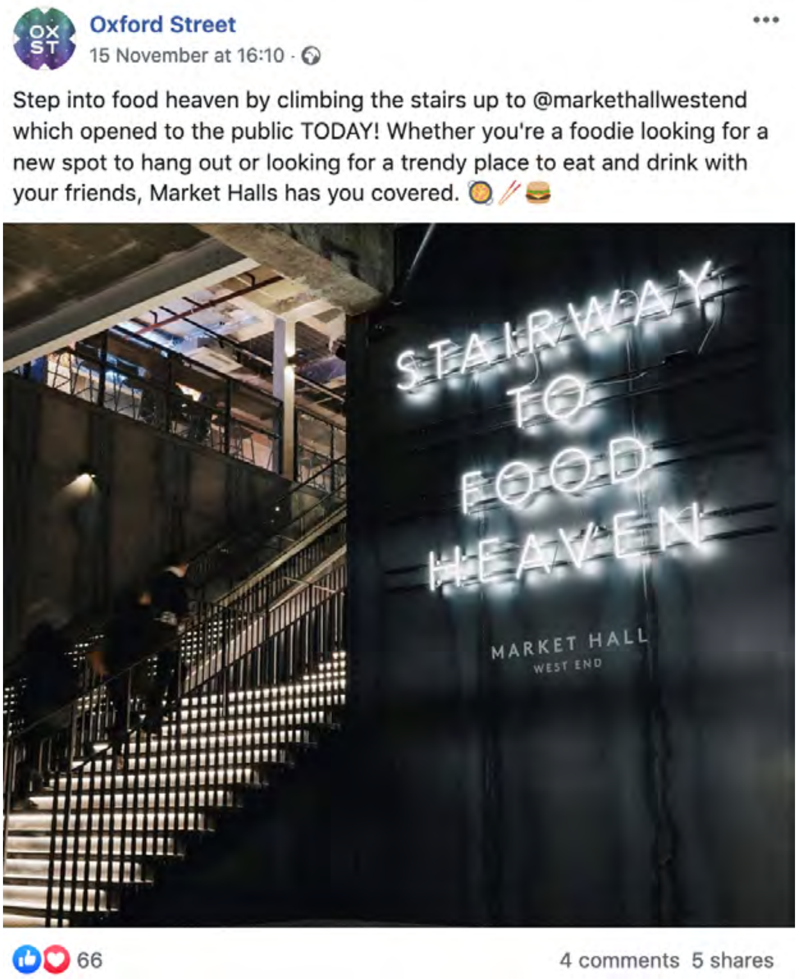


Figure B.36: An example Facebook post from Oxford Street, London. Source: https://www.facebook.com/OxfordStW1

.

Figure B.37: Regent Street, London. Source: http://www.regentstreetonline.com

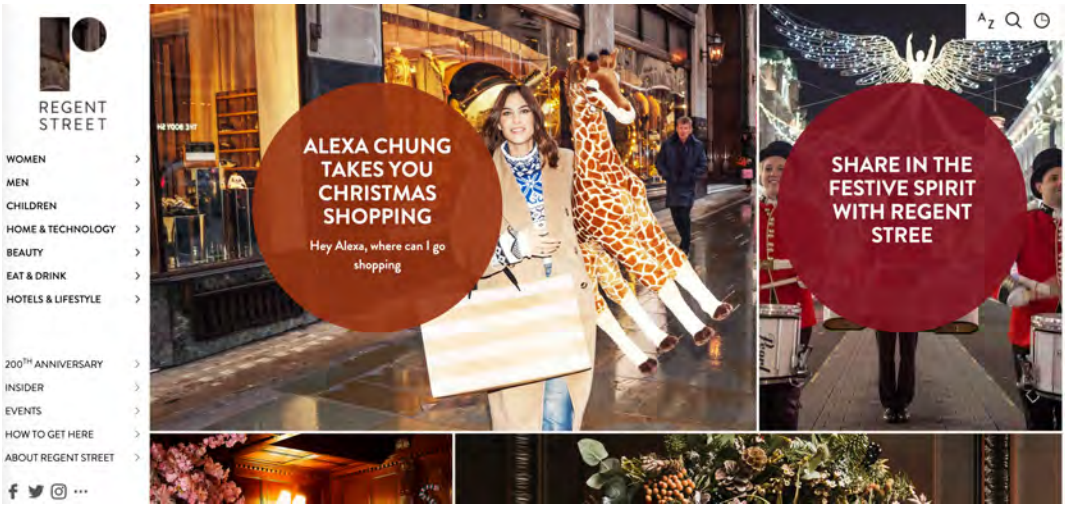


Figure B.38: Regent Steet, London's web presence. Source: https://www.regentstreetonline.com

Regent Street maintains an online presence, with a website (see Figure B.38) and accounts on various social media platforms - such as Facebook, Twitter, Instagram, Pinterest, YouTube, and Weibo. Example social media posts can be seen in Figures B.39 and B.40.

B.11 Seven Dials, London

Seven Dials (see Figure B.41), the name now given to an area of streets surrounding a road junction featuring a six-faced sun dial (the 7th being the monument itself), is now home to approximately 100 retailers and an additional 90 cafes, bars, and restaurants. The entire area is predominantly owned by a single company. As well as a main website (see Figure B.42), Seven Dials maintains a presence on multiple social media platforms, actively promoting Facebook, Twitter, Instagram, and YouTube. Example posts can be seen in Figures B.43 and B.44.

Figure B.39: An example Twitter post from Regent Street, London. Source: https://twitter.com/RegentStreetW1.

Figure B.40: An example Instagram post from Regent Street, London. Source: https://www.instagram.com/RegentStreetW1

.

Figure B.41: Seven Dials, London. Source: http://www.facebook.com/sevendials

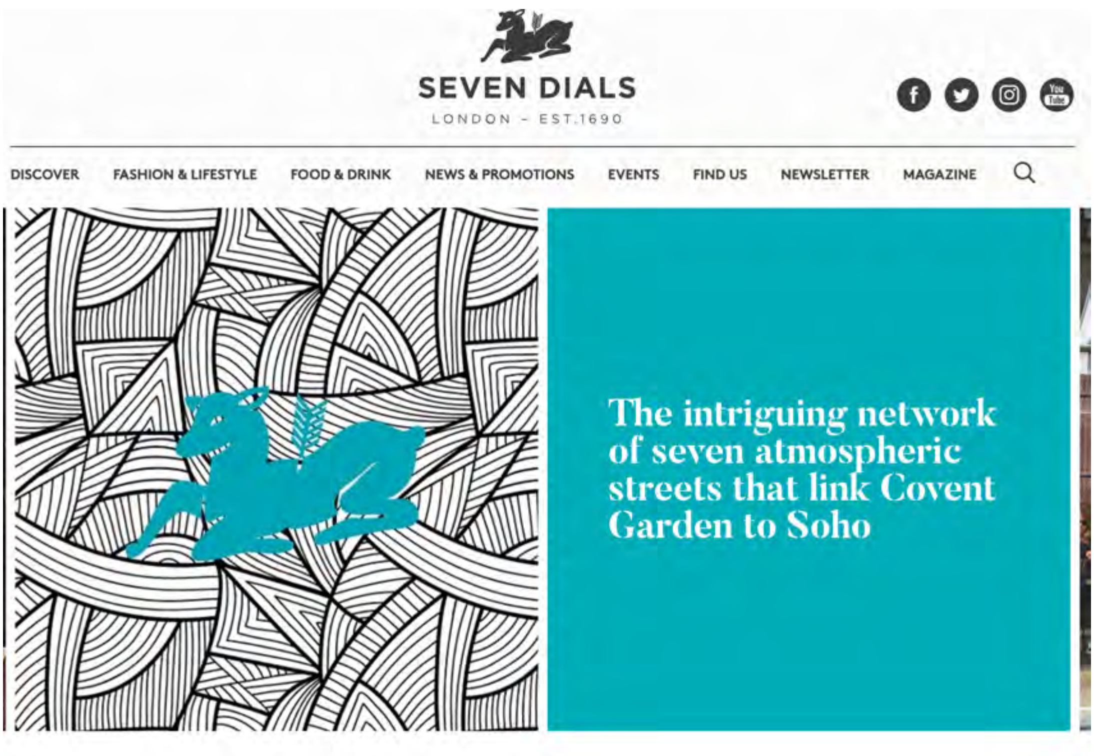


Figure B.42: Seven Dials, London's web presence. Source: https://www.sevendials.co.uk

Figure B.43: An example Twitter post from Seven Dials, London. Source: https://twitter.com/7dialsLondon.

Figure B.44: An example Facebook post from Seven Dials, London. Source: https://www.facebook.com/sevendials

.

Figure B.45: Trinity Leeds. Source: http://www.facebook.com/trinityleeds

Figure B.46: Trinity Leeds' web presence. Source: https://www.trinityleeds.co.uk

B.12 Trinity Leeds, Leeds

The newest retail location included in this thesis, Trinity Leeds (shown in Figure B.45) is a city-centre shopping and leisure venue that first opened in March 2013. With over 120 tenant retailers and services, Trinity Leeds covers over 90,000m^2 of total retail floor space. Maintaining both a website (Figure B.46), and social media presence on various platforms, Trinity Leeds actively promotes various events and partnerships - including with local rugby team, Leeds Rhinos. Example social media posts can be seen in Figures B.47 and B.48.

Figure B.47: An example Twitter post from Trinity Leeds. Source: https://twitter.com/trinityleeds.

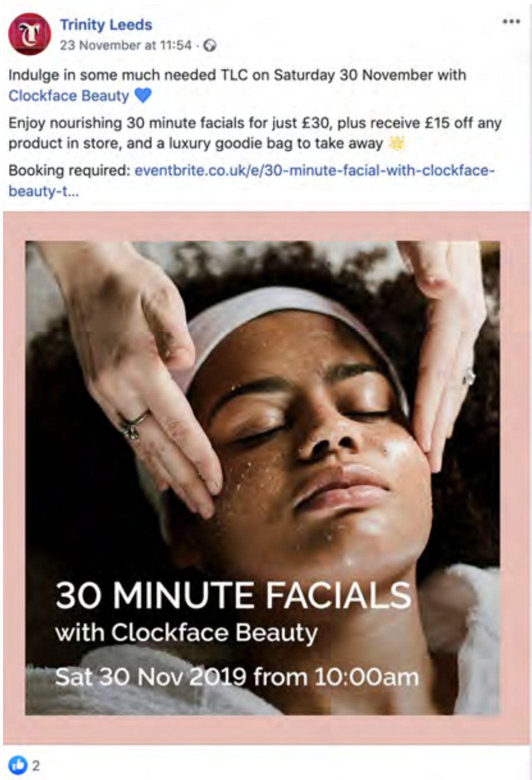


Figure B.48: An example Facebook post from Trinity Leeds. Source: https://www.facebook.com/trinityleeds

.

Appendix C

Network Churn - Hypotheses Results Tables

This appendix includes tables that contain the results of logistic regression tests conducted as part of the study presented in Chapter 6.

C.1 Hypothesis 1C

Table C.1: Hypothesis 1C - Logistic regression results table - Regent Street, London.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
1	0.3132	0.2840	1712.3	1711.0	1715.0	0.000799
2	-0.1820	0.2123	2592.2	2591.5	2595.5	0.000318
3	2.0859	0.5788	862.5	846.6	850.6	0.019267
4	0.8275	0.3111	1703.3	1695.7	1699.7	0.004862
5	0.2209	0.2075	2982.3	2981.1	2985.1	0.000443
6	0.2655	0.3224	1545.5	1544.8	1548.8	0.000480
7	0.6612	0.2967	1994.8	1989.5	1993.5	0.002882
8	0.5796	0.2593	2331.6	2326.3	2330.3	0.002463
9	0.2363	0.2957	1764.3	1763.6	1767.6	0.000394
Mean	0.5565	0.3075	1943.2	1938.9	1942.9	0.003545

Table C.2: Hypothesis 1C - Logistic regression results table - Carnaby London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	-0.0577	0.2503	2384.6	2384.5	2388.5	0.000024
2	1.2640	0.2413	3613.4	3584.2	3588.2	0.009666
3	1.0773	0.2665	2940.4	2922.2	2926.2	0.006755
4	0.8097	0.2786	2581.2	2572.1	2376.1	0.003830
5	0.3636	0.1820	4679.4	4675.3	4679.3	0.001016
6	0.9283	0.3330	2031.9	2023.4	2027.4	0.004450
7	0.4986	0.2665	2617.6	2614.0	2618.0	0.001515
8	0.7265	0.2886	2497.6	2490.7	2494.7	0.002970
9	0.5502	0.2522	2936.5	2931.5	2935.5	0.001857
Mean	0.6845	0.2621	2920.5	2910.9	2914.9	0.003565

Table C.3: Hypothesis 1C - Logistic regression results table - Seven Dials, London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	0.8571	0.4822	1046.5	1043.1	1047.1	0.003549
2	-0.4440	0.4081	1003.3	1002.2	1006.2	0.001205
3	0.6532	0.4415	1243.7	1241.4	1245.4	0.002036
4	0.2843	0.3453	1633.3	1632.6	1636.6	0.000472
5	0.2217	0.2689	2181.4	2180.8	2184.8	0.000318
6	0.6161	0.4390	1202.7	1200.5	1204.5	0.001872
7	1.4357	0.4306	1512.5	1499.6	1503.6	0.009266
8	-0.1388	0.2758	2122.2	2121.9	2125.9	0.000132
9	1.5550	0.5529	1111.2	1101.8	1105.8	0.008971
Mean	0.5608	0.4049	1450.8	1447.1	1451.1	0.003091

Table C.4: Hypothesis 1C - Logistic regression results table - Oxford Street.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	0.1103	0.2469	1838.1	1837.9	1841.9	0.000119
2	-0.6120	0.2531	1645.5	1640.0	1644.0	0.003679
3	-0.3458	0.2802	1491.7	1490.2	1494.2	0.001062
4	0.5124	0.2895	1618.0	1614.8	1618.8	0.002152
5	0.1166	0.1675	3689.5	3689.0	3693.0	0.000153
6	0.6627	0.3034	1622.0	1616.9	1620.9	0.003313
7	0.8024	0.3332	1408.2	1402.1	1406.1	0.004638
8	1.2403	0.1619	5026.6	4961.4	4965.4	0.015289
9	-0.0586	0.2443	2052.2	2052.2	2056.2	0.000030
Mean	0.2698	0.2533	2265.8	2256.1	2260.1	0.003381

Table C.5: Hypothesis 1C - Logistic regression results table - Covent Garden, London.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
1	1.0948	0.2730	2794.5	2776.9	2780.9	0.006663
2	0.1052	0.2220	3477.6	3477.4	3481.4	0.000070
3	0.0986	0.2495	2981.5	2981.3	2985.3	0.000056
4	-0.0840	0.1470	6784.6	6784.3	6788.3	0.000054
5	0.4936	0.2462	3316.8	3312.6	3316.6	0.0013450
6	0.8440	0.3100	2338.4	2330.4	2334.4	0.003561
7	0.2543	0.1659	5845.3	5842.9	5846.9	0.000448
8	0.8148	0.2609	3229.4	3218.9	3322.9	0.003417
9	0.3944	0.2472	3140.6	3137.9	3141.9	0.000877
Mean	0.4462	0.2357	3767.6	3762.5	3777.6	0.001832

Table C.6: Hypothesis 1C - Logistic regression results table - London West End.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
1	0.1586	0.1719	5461.0	5460.1	5464.1	0.000171
2	0.7690	0.1883	5556.8	5538.8	5542.8	0.003449
3	0.8072	0.1943	5398.0	5379.5	5383.5	0.003681
4	0.6258	0.1563	7342.4	7325.5	7329.5	0.002526
5	0.5944	0.2610	3252.5	3247.1	3251.1	0.001755
6	0.1578	0.2225	3778.9	3778.4	3782.4	0.000142
7	0.5639	0.1845	5477.1	5467.3	5471.3	0.001915
8	0.7502	0.2877	2701.2	2694.0	2698.0	0.002778
9	0.2481	0.2573	2961.8	2960.9	2964.9	0.000333
Mean	0.5194	0.2183	4658.9	4650.2	4654.4	0.001861

C.2 Hypothesis 1D

Table C.7: Hypothesis 1D - Logistic regression results table - Regent Street, London.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
1	6.2542	4.4858	1712.3	1709.9	1713.9	0.001520
2	12.7790	4.4464	2592.2	2581.1	2585.1	0.004891
3	5.6210	7.4700	862.5	861.9	865.9	0.000809
4	-2.6820	3.6250	1703.3	1702.9	1706.9	0.000313
5	-1.2623	2.3008	2982.3	2982.0	2986.0	0.000109
6	10.5835	6.0509	1545.5	1541.5	1545.5	0.002763
7	-3.4005	2.6864	1994.8	1993.4	1997.4	0.000762
8	5.9235	4.1099	2331.6	2329.1	2333.1	0.001178
9	16.9884	7.0464	1764.3	1756.1	1760.1	0.004947
Mean	5.6450	4.6913	1943.2	1939.8	1943.8	0.001921

Table C.8: Hypothesis 1D - Logistic regression results table - Carnaby London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	12.4827	5.5094	2384.6	2377.7	2381.7	0.003154
2	-1.148	2.4580	3615.4	3615.2	3619.2	0.000065
3	6.8182	4.0816	2940.4	2936.9	2940.9	0.001270
4	-3.5919	2.4836	2581.2	2579.4	2583.4	0.000740
5	-2.6221	2.0928	4679.4	4678.0	4682.0	0.000350
6	4.3869	4.5800	2031.9	2030.8	2034.8	0.000557
7	-3.0099	2.8119	2617.6	2616.6	2620.6	0.000416
8	6.0161	4.4597	2497.6	2495.5	2499.5	0.000938
9	-3.5076	2.4289	2936.5	2934.7	2938.7	0.000662
Mean	1.7583	4.4340	2920.5	2918.3	2922.3	0.000906

Table C.9: Hypothesis 1D - Logistic regression results table - Seven Dials, London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	-2.9352	3.4657	1046.5	1045.9	1049.9	0.000648
2	2.5014	5.0321	1003.3	1003.1	1007.1	0.000286
3	-0.0597	4.1933	1243.7	1243.7	1247.7	0.000001
4	1.4983	3.8852	1633.2	1633.2	1637.2	0.000105
5	3.2851	3.4973	2181.4	2180.4	2184.4	0.000504
6	-0.5734	4.0052	1202.7	1202.6	1206.6	0.000018
7	-5.4390	2.2690	1512.5	1508.0	1512.0	0.003208
8	6.2768	3.8489	2122.2	2119.0	2123.0	0.001681
9	6.9773	6.2290	1111.2	1109.7	1113.7	0.001431
Mean	1.2813	4.0473	1450.7	1449.5	1453.5	0.000876

Table C.10: Hypothesis 1D - Logistic regression results table - Oxford Street, London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	-3.7455	3.2810	1838.1	1836.9	1840.9	0.00660
2	18.8567	8.3478	1645.6	1638.4	1642.4	0.004691
3	32.4764	11.2162	1491.7	1478.5	1482.5	0.009401
4	1.7340	4.8059	1618.0	1617.9	1621.9	0.000091
5	-0.1148	2.6267	3689.5	3689.5	3693.5	0.000001
6	-0.6979	4.1713	1622.0	1621.9	1625.9	0.000018
7	-1.8751	4.0007	1408.2	1408.0	1412.0	0.000151
8	29.1952	5.1299	5026.6	4977.6	4981.6	0.011490
9	13.7192	6.8909	2052.2	2046.9	2050.9	0.002754
Mean	9.9498	5.6078	2265.8	2257.3	2261.3	0.003251

Table C.11: Hypothesis 1D - Logistic regression results table - Covent Garden, London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	17.4684	6.5148	2794.5	2784.8	2788.8	0.003673
2	11.5058	5.1216	3477.6	3471.1	3475.1	0.002008
3	24.1162	7.4427	2981.5	2966.4	2970.4	0.005356
4	7.9107	3.1869	6784.6	6777.1	6781.1	0.001261
5	6.9964	4.6701	3316.8	3314.1	3318.1	0.000875
6	22.1181	8.4035	2338.4	2328.4	2332.4	0.004443
7	7.0286	3.4654	5845.3	5840.3	5844.3	0.000937
8	39.4323	10.0170	3229.4	3205.2	3209.2	0.007872
9	19.0650	7.2639	3140.6	3131.0	3135.0	0.003195
Mean	17.2935	6.2308	3767.6	3757.6	3761.6	0.003291

Table C.12: Hypothesis 1D - Logistic regression results table - London West End.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	12.3291	4.1860	5461.0	5449.5	5453.5	0.002275
2	30.0108	6.1170	5556.6	5522.0	5526.0	0.006713
3	23.7281	5.6794	5398.0	5373.9	5377.9	0.004808
4	73.3921	7.2326	7342.4	7175.8	7179.8	0.024921
5	35.0549	9.0670	3252.5	3230.2	3234.2	0.007158
6	15.1334	6.2601	3778.9	3771.1	3775.1	0.002155
7	42.6921	7.4803	5477.1	5429.9	5433.9	0.009233
8	22.0666	8.8188	2701.2	2692.6	2696.6	0.003293
9	11.6960	6.7894	2961.8	2958.0	2962.0	0.001345
Mean	29.6781	6.8478	4658.8	4622.6	4626.6	0.006878

C.3 Hypothesis 4

Table C.13: Hypothesis 4 - Logistic regression results table - Regent Street, London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	0.2927	0.2822	1712.3	1711.3	1715.	0.000638
2	0.5445	0.2004	2592.2	2585.8	2589.8	0.002834
3	0.9105	0.3184	862.5	855.8	859.8	0.008064
4	0.0891	0.2582	1703.3	1703.2	1707.2	0.000074
5	-0.4005	0.2590	2982.3	2979.6	2983.6	0.001030
6	-0.2857	0.3656	1545.5	1544.8	1548.8	0.000460
7	0.1437	0.2497	1994.8	1994.5	1998.5	0.000174
8	0.4921	0.2146	2331.6	2326.9	2330.9	0.002189
9	0.9909	0.2134	1764.3	1746.9	1750.9	0.010513
Mean	0.3086	0.2624	1943.2	1938.8	1942.8	0.002886

Table C.14: Hypothesis 4 - Logistic regression results table - Carnaby London.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
1	0.2677	0.2479	2384.6	2383.5	2387.5	0.000494
2	0.0399	0.2057	3615.4	3615.4	3619.4	0.000012
3	0.9714	0.1718	2940.4	2914.7	2918.7	0.009561
4	0.1518	0.2607	2581.2	2580.9	2584.9	0.000136
5	-0.0394	0.1901	4679.4	4679.4	4683.4	0.000011
6	0.7872	0.2394	2031.9	2023.0	2027.0	0.004643
7	0.4868	0.2212	2617.6	2613.4	2617.4	0.001759
8	0.5749	0.2178	2497.6	2491.6	2495.6	0.002590
9	0.3276	0.2198	2936.5	2934.5	2938.5	0.000752
Mean	0.3964	0.2194	2870.5	2915.2	2919.2	0.002218

Table C.15: Hypothesis 4 - Logistic regression results table - Seven Dials, London.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
1	0.5793	0.3373	1046.5	1044.0	1048.0	0.002614
2	-0.6798	0.5887	1003.3	1001.7	1005.7	0.001761
3	0.1163	0.3928	1243.7	1243.6	1247.6	0.000074
4	-0.0054	0.3462	1633.3	1633.3	1637.3	0.000001
5	-0.4797	0.3630	2181.4	2179.3	2183.3	0.001054
6	-0.6493	0.5871	1202.7	1200.9	1204.9	0.001556
7	-0.2675	0.4195	1512.5	1512.0	1516.0	0.000317
8	0.2240	0.2713	2122.5	2121.5	2125.5	0.000339
9	0.1689	0.3943	1111.2	1111.0	1115.0	0.000167
Mean	-0.1104	0.4111	1450.8	1449.7	1453.7	0.000876

Table C.16: Hypothesis 4 - Logistic regression results table - Oxford Street, London.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
1	0.1442	0.1922	1838.1	1837.5	1841.5	0.000322
2	0.0021	0.2276	1645.6	1645.6	1649.6	0.000001
3	0.2580	0.2229	1491.7	1490.4	1494.4	0.000904
4	0.2216	0.2173	1618.0	1617.0	1621.0	0.000655
5	0.2337	0.1306	3689.5	3686.5	3690.5	0.000957
6	0.2832	0.2174	1622.0	1620.4	1624.4	0.001048
7	0.0774	0.2561	1408.2	1408.2	1412.2	0.000067
8	1.2634	0.0879	5026.6	4848.8	4849.8	0.042201
9	0.2906	0.1930	2052.2	2050.1	2054.1	0.001106
Mean	0.3082	0.1939	2265.8	2244.9	2248.6	0.005251

Table C.17: Hypothesis 4 - Logistic regression results table - Covent Garden, London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	0.8910	0.1939	2794.5	2777.4	2781.4	0.006453
2	0.8390	0.1791	3477.6	3459.7	3463.7	0.005503
3	0.7977	0.2050	2981.5	2969.1	2973.1	0.004394
4	0.3968	0.1442	6784.6	6777.8	6781.8	0.001142
5	0.8560	0.1818	3316.8	3298.8	3302.8	0.005759
6	1.0778	0.2090	2338.4	2317.7	2321.7	0.009208
7	0.5939	0.1384	5845.3	5829.4	5833.4	0.003004
8	0.8152	0.1845	3229.4	3213.4	3217.4	0.005201
9	1.3545	0.1573	3140.6	3084.6	3088.6	0.018639
Mean	0.8469	0.1770	3767.6	3747.5	3751.5	0.006589

Table C.18: Hypothesis 4 - Logistic regression results table - London West End.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
1	0.4951	0.1593	5461.0	5452.5	5456.5	0.001680
2	0.6320	0.1419	5556.6	5539.6	5543.6	0.003286
3	0.7584	0.1424	5398.0	5374.5	5378.5	0.004658
4	1.7342	0.0865	7342.4	7041.0	7045.0	0.045000
5	0.7093	0.1893	3252.5	3240.8	3244.8	0.003774
6	1.0542	0.1592	3778.9	3744.5	3748.5	0.009549
7	1.0043	0.1290	5477.1	5428.9	5432.9	0.009424
8	1.2304	0.1817	2701.2	2666.3	2670.3	0.013835
9	1.0243	0.1840	2961.8	2937.4	2941.4	0.008552
Mean	0.9602	0.1526	4658.8	4602.8	4606.8	0.011037

C.4 Hypothesis 7D

Table C.19: Hypothesis 7D - Logistic regression results table - Regent Street, London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
2	0.2400	9.2056	1713.4	1713.4	1717.4	0.372162
3	18.2493	11.1102	507.6	506.0	501.0	0.424202
4	-7.2036	10.4521	1548.1	1547.6	1551.6	0.098318
5	23.0946	6.6357	2654.1	2643.1	2647.1	0.128562
6	11.5409	6.7237	1007.4	1005.7	1009.7	0.366111
7	5.7082	10.2503	1533.1	1532.8	1536.8	0.247590
8	21.8502	7.5506	1823.5	1817.8	1821.8	0.237943
9	9.3082	7.5167	1208.0	1207.1	1211.1	0.331514
Mean	10.3485	8.6806	1499.4	1496.7	1499.6	0.275800

Table C.20: Hypothesis 7D - Logistic regression results table - Carnaby London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
2	9.3255	8.3085	1938.5	1937.6	1941.6	0.495809
3	72.0388	10.2469	2429.4	2379.2	2383.2	0.205969
4	25.7515	8.2758	2246.3	2240.1	2244.1	0.141752
5	3.9893	5.9706	4069.0	4068.6	4072.6	0.147784
6	20.8270	6.0937	1538.3	1530.4	1534.4	0.258787
7	5.8832	6.1285	2319.5	2318.9	2322.9	0.122160
8	22.9960	5.2870	2197.8	2185.9	2189.9	0.133037
9	13.9547	5.4262	2456.9	2452.7	2456.7	0.176797
Mean	21.8458	6.9672	2399.5	2389.2	2393.2	0.210262

Table C.21: Hypothesis 7D - Logistic regression results table - Seven Dials.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
2	3.1227	12.5227	702.6	702.5	706.5	0.314719
3	21.0924	6.6764	659.9	654.3	658.3	0.494676
4	6.9466	13.5047	1445.4	1445.1	1449.1	0.125765
5	6.6711	8.1893	1933.3	1932.8	1936.8	0.128050
6	20.9696	9.7866	851.0	848.0	852.0	0.310556
7	6.1768	18.1419	844.8	844.7	848.7	0.462787
8	9.1237	7.1189	1644.6	1643.4	1647.4	0.246778
9	12.8201	7.9799	720.7	719.1	723.1	0.367056
Mean	10.8654	10.4901	1100.3	1098.7	1102.7	0.306209

Table C.22: Hypothesis 7D - Logistic regression results table - Oxford Street, London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
2	15.1331	9.7121	1215.3	1213.9	1217.9	0.276884
3	23.5182	9.5025	1366.9	1363.2	1367.2	0.091237
4	1.4471	14.7908	1423.3	1423.3	1427.3	0.127682
5	3.5046	9.7645	3279.3	3279.1	3283.1	0.126942
6	10.1606	5.8223	1232.2	1230.4	1234.4	0.253947
7	11.5350	7.1200	1137.3	1135.8	1139.8	0.202296
8	15.9466	5.2545	2548.2	2541.7	2545.7	0.543351
9	11.4462	6.9762	1129.4	1126.1	1130.1	0.469357
Mean	11.5864	8.6179	1666.5	1664.2	1668.2	0.261462

Table C.23: Hypothesis 7D - Logistic regression results table - Covent Garden, London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
2	8.4997	2.6816	2924.9	2918.6	2922.6	0.170662
3	14.4222	3.7793	2505.6	2496.3	2500.3	0.171058
4	8.8089	4.0977	6258.9	6255.5	6259.3	0.087883
5	9.5786	4.5519	2802.9	2800.2	2804.2	0.164394
6	28.9642	4.6539	1882.0	1860.6	1864.6	0.211553
7	6.8552	4.9991	4790.3	4789.0	4793.0	0.196901
8	11.1426	3.8847	1698.5	1694.5	1698.5	0.489716
9	42.0807	9.6585	2700.0	2674.2	2678.2	0.154630
Mean	16.2940	4.7883	3195.4	3186.1	3190.1	0.205850

Table C.24: Hypothesis 7D - Logistic regression results table - London West End.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
2	14.2555	2.8645	3220.9	3205.6	3209.6	0.443315
3	20.0748	3.3029	4854.5	4824.3	4828.3	0.113477
4	90.7236	6.9038	6653.5	6401.2	6405.2	0.139626
5	14.6522	3.2105	2692.1	2679.1	2683.1	0.182767
6	23.8597	3.9516	3395.5	3369.4	3373.4	0.113240
7	14.5898	2.9212	4979.0	4961.2	4965.2	0.100438
8	21.2677	4.5067	2342.9	2329.5	2333.5	0.141845
9	21.8008	4.6411	2667.6	2649.6	2653.6	0.109002
Mean	27.6530	4.0378	3850.8	3802.5	3806.5	0.167964

C.5 Hypothesis 9

Table C.25: Hypothesis 9 - Logistic regression results table - Regent Street, London.

#	Estimate	Std. Error	Null Dev.	Residual Dev.	AIC	R^2
2	0.0068	0.0124	1713.4	1713.1	1717.1	0.372261
3	0.1192	0.0297	507.6	497.6	501.6	0.434045
4	0.0125	0.0064	1548.1	1545.6	1549.6	0.099571
5	0.0224	0.0084	2654.1	2647.1	2651.1	0.127051
6	0.0854	0.0170	1007.4	988.0	992.0	0.377783
7	0.0040	0.0032	1533.1	1532.2	1536.2	0.247916
8	0.0722	0.0184	1823.5	1810.9	1814.9	0.241044
9	0.1641	0.0222	1208.0	1152.6	1156.6	0.363158
Mean	0.0608	0.0147	1499.4	1485.9	1489.9	0.282854

Table C.26: Hypothesis 9 - Logistic regression results table - Carnaby London.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
2	0.0358	0.0088	1938.5	1925.5	1929.5	0.499185
3	0.0849	0.0092	2429.4	2262.7	2266.7	0.247774
4	0.0824	0.0161	2246.3	2212.7	2216.7	0.153028
5	0.0036	0.0031	4069.0	4068.1	4072.1	0.147906
6	0.0695	0.0097	1538.3	1488.1	1492.1	0.280205
7	0.0571	0.0117	2319.5	2297.2	2301.2	0.130935
8	0.1055	0.0101	2197.8	2073.9	2077.9	0.180277
9	0.0292	0.0088	2456.9	2440.5	2444.5	0.181182
Mean	0.0585	0.0097	2399.5	2346.1	2350.1	0.227561

Table C.27: Hypothesis 9 - Logistic regression results table - Seven Dials, London.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
2	0.0383	0.0195	702.6	700.1	704.1	0.317124
3	0.1704	0.0265	659.9	598.8	602.8	0.539275
4	0.0487	0.0262	1445.4	1442.4	1446.4	0.127559
5	0.0071	0.0219	1933.3	1933.2	1937.2	0.127828
6	0.0278	0.0155	851.0	848.8	852.8	0.309901
7	0.1207	0.0232	844.8	825.6	829.6	0.475533
8	0.0404	0.0148	1644.6	1638.8	1642.8	0.249079
9	0.0929	0.0171	720.7	686.1	690.1	0.397191
Mean	0.0683	0.0206	1100.3	1084.2	1088.2	0.317936

Table C.28: Hypothesis 9 - Logistic regression results table - Oxford Street, London.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
2	0.020	0.0080	1215.3	1210.8	1214.8	0.278810
3	0.094	0.0256	1366.9	1347.8	1351.8	0.102043
4	-0.0136	0.0202	1423.3	1423.0	1427.0	0.127884
5	0.0535	0.0147	3279.3	3264.9	3268.9	0.131253
6	0.1040	0.0213	1232.2	1210.0	1214.0	0.266943
7	0.1872	0.0292	1137.3	1098.7	1102.7	0.229507
8	0.1928	0.0223	2548.2	2448.2	2452.2	0.561804
9	0.0833	0.0144	1129.4	1102.8	1106.8	0.480771
Mean	0.0902	0.0195	1666.5	1638.3	1642.3	0.272377

Table C.29: Hypothesis 9 - Logistic regression results table - Covent Garden, London.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
2	0.1010	0.0120	2924.9	2813.9	2817.	0.202370
3	0.1703	0.0123	2505.6	2265.7	2269.7	0.251222
4	0.1062	0.0111	6258.9	6161.9	6165.9	0.103243
5	0.0420	0.0077	2802.9	2770.6	2774.6	0.173710
6	0.0935	0.0110	1882.0	1761.2	1765.2	0.255105
7	0.1290	0.0117	4790.3	4636.0	4640.0	0.224800
8	0.0716	0.0111	1698.5	1633.7	1637.7	0.508569
9	0.2264	0.0149	2700.0	2285.2	2289.2	0.281932
Mean	0.1175	0.0115	3195.4	3041.0	3045.0	0.250118

Table C.30: Hypothesis 9 - Logistic regression results table - London West End.

#	**Estimate**	**Std. Error**	**Null Dev.**	**Residual Dev.**	**AIC**	R^2
2	0.1500	0.0099	3220.9	2886.5	2890.5	0.501131
3	0.1473	0.0071	4854.5	4328.3	4332.3	0.210148
4	0.1778	0.0064	6653.5	5313.4	5317.4	0.296610
5	0.1804	0.0123	2692.1	2354.9	2358.9	0.284855
6	0.2488	0.0113	3395.5	2675.5	2679.5	0.302353
7	0.0870	0.0057	4979.0	4677.8	4681.8	0.155040
8	0.1701	0.0120	2342.9	2013.0	2017.0	0.261510
9	0.2250	0.0138	2667.6	2215.0	2219.0	0.259299
Mean	0.1733	0.0098	3850.8	3308.1	3312.1	0.283869

C.6 Hypothesis 11

Table C.31: Hypothesis 11 - Logistic regression results table - Regent Street, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1713.4	1713.1	1719.1	0.372263
3	507.6	497.3	503.3	0.434428
4	1548.1	543.4	1549.4	0.100956
5	2654.1	2639.3	2645.3	0.129957
6	1007.4	987.8	993.8	0.377928
7	1533.1	1532.1	1538.1	0.247977
8	1823.5	1807.1	1813.1	0.242770
9	1208.0	1151.5	1157.5	0.363817
Mean	1499.4	1359.0	1490.0	0.283762

Table C.32: Hypothesis 11 - Logistic regression results table - Carnaby London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1938.6	1295.5	1931.5	0.499202
3	2429.4	2270.3	2276.3	0.245065
4	2246.3	2210.9	2216.9	0.153774
5	4069.0	4067.9	4073.9	0.147960
6	1538.3	1486.0	1492.1	0.281278
7	2319.5	2297.2	2303.2	0.130938
8	2197.8	2072.0	2078.0	0.181065
9	2456.9	2438.2	2444.2	0.181997
Mean	2399.5	2267.3	2352.0	0.227660

Table C.33: Hypothesis 4 - Logistic regression results table - Seven Dials, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	702.5	700.0	706.0	0.317235
3	659.9	598.7	604.7	0.539326
4	1445.4	1442.4	1448.4	0.127577
5	1933.3	1932.7	1938.7	0.128077
6	851.0	847.0	853.0	0.311393
7	844.8	825.5	831.5	0.475593
8	1644.6	1638.2	1644.2	0.249368
9	720.6	685.4	691.4	0.397831
Mean	1100.3	1083.7	1089.7	0.318300

Table C.34: Hypothesis 11 - Logistic regression results table - Oxford Street, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1215.4	1201.1	1216.1	0.279282
3	1366.9	1347.3	1253.3	0.102431
4	1423.3	1422.9	1428.9	0.127928
5	3279.3	3263.0	3269.0	0.131851
6	1232.2	1209.3	1215.3	0.267382
7	1137.3	1098.3	1104.3	0.229800
8	2548.2	2448.0	2454.0	0.561856
9	1129.4	1101.6	1107.6	0.481358
Mean	1666.5	1636.4	1631.1	0.272736

Table C.35: Hypothesis 11 - Logistic regression results table - Covent Garden, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	2924.9	2808.4	2814.8	0.203832
3	2505.6	2265.4	2271.4	0.251306
4	6258.9	6161.9	6167.9	0.103243
5	2802.7	2770.5	2776.5	0.173748
6	1882.0	1758.2	1764.2	0.256409
7	4790.3	4633.4	4639.4	0.225281
8	1698.5	1633.7	1639.7	0.508574
9	2700.0	2282.8	2288.8	0.282699
Mean	3195.4	3039.3	3045.3	0.250637

Table C.36: Hypothesis 11 - Logistic regression results table - London West End.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	3220.9	2883.4	2889.4	0.501685
3	4854.5	4303.3	4309.3	0.214989
4	6653.5	5252.1	5258.1	0.305321
5	2692.1	2354.7	2360.7	0.284905
6	3395.5	2669.5	2675.5	0.303966
7	4979.0	4677.8	4683.8	0.155041
8	2342.9	2010.2	2016.2	0.262567
9	2667.6	2209.7	2215.7	0.261131
Mean	3850.8	3295.1	3301.1	0.286201

Appendix D

Social Media Engagement & Network Churn - Hypotheses Results Tables

D.1 Hypothesis 3

D.1.1 Retweet Count

Regent Street, London

Table D.1: Hypothesis 3 - Retweet Count - Logistic regression results table - Regent Street, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1713.40	1713.00	1721.00	0.372314
3	507.59	496.59	504.59	0.435217
4	1548.10	1541.40	1549.40	0.102167
5	2654.10	2632.80	2640.80	0.132374
6	1007.43	987.13	995.13	0.378358
7	1533.10	1531.60	1539.60	0.248258
8	1823.50	1805.40	1813.40	0.243554
9	1208.00	1149.70	1157.70	0.364818
Mean	1499.40	1482.20	1490.20	0.284632

Carnaby London

Table D.2: Hypothesis 3 - Retweet Count - Logistic regression results table - Carnaby London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1938.50	1924.70	1932.70	0.499399
3	2429.40	2265.30	2273.70	0.246865
4	2246.30	2210.10	2218.10	0.154101
5	4069.00	4064.30	4072.30	0.148797
6	1538.30	1481.20	1489.20	0.283740
7	2319.50	2295.50	2303.50	0.131618
8	2197.80	2067.30	2075.30	0.183038
9	2456.90	2435.60	2443.60	0.182925
Mean	2399.46	2343.00	2351.05	0.228810

Seven Dials, London

Table D.3: Hypothesis 3 - Retweet Count - Logistic regression results table - Seven Dials, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	702.55	699.86	707.86	0.317414
3	659.85	598.54	606.54	0.539462
4	1445.40	1436.30	1444.30	0.131603
5	1933.30	1928.60	1936.60	0.130186
6	851.03	846.65	864.70	0.311697
7	844.83	824.77	832.77	0.476072
8	1644.60	1638.20	1646.20	0.249388
9	720.65	685.15	693.15	0.398084
Mean	1100.28	1082.26	1090.26	0.319238

Oxford Street, London

Table D.4: Hypothesis 3 - Retweet Count - Logistic regression results table - Oxford Street, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1215.30	1209.60	1217.60	0.279585
3	1366.90	1346.90	1354.90	0.102693
4	1423.30	1421.70	1429.70	0.128694
5	3279.30	3263.30	3271.00	0.131856
6	1232.20	1209.00	1217.00	0.267572
7	1137.30	1097.80	1105.80	0.230147
8	2548.20	2446.90	2454.90	0.562057
9	1129.40	1098.90	1106.90	0.482652
Mean	1722.74	1636.73	1644.73	0.273157

Covent Garden, London

Table D.5: Hypothesis 3 - Retweet Count - Logistic regression results table - Covent Garden, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	2924.90	2808.40	2816.40	0.203832
3	2505.60	2264.10	2272.10	0.251757
4	6258.90	6159.60	6167.60	0.103621
5	2802.90	2770.00	2778.00	0.173906
6	1882.00	1746.50	1754.50	0.261529
7	4790.30	4633.00	4641.00	0.225359
8	1698.50	1633.00	1641.00	0.508780
9	2700.00	2281.70	2289.70	0.283066
Mean	3195.39	3037.04	3045.04	0.251481

London West End

Table D.6: Hypothesis 3 - Retweet Count - Logistic regression results table - London West End.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	3220.90	2883.40	2891.40	0.510689
3	4854.50	4303.30	4311.30	0.214994
4	6653.50	5252.10	5260.10	0.305321
5	2692.10	2354.70	2362.70	0.284906
6	3395.50	2669.50	2677.50	0.303967
7	4979.00	4677.30	4685.30	0.155136
8	2342.90	2009.80	2017.80	0.262691
9	2667.60	2209.50	2217.50	0.261211
Mean	3850.75	3294.95	3302.95	0.286239

D.1.2 Retweeted

Regent Street, London

Table D.7: Hypothesis 3 - Retweeted - Logistic regression results table - Regent Street, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1713.40	1713.10	1721.10	0.372264
3	507.59	496.59	504.59	0.435217
4	1548.10	1541.40	1549.40	0.102167
5	2654.10	2639.10	2647.10	0.130033
6	1007.43	987.13	995.13	0.378358
7	1533.10	1530.00	1538.00	0.249058
8	1823.50	1805.40	1813.40	0.243554
9	1208.00	1149.70	1157.70	0.364818
Mean	1499.40	1482.80	1490.80	0.284434

Carnaby London

Table D.8: Hypothesis 3 - Retweeted - Logistic regression results table - Carnaby London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1938.50	1924.70	1932.70	0.499399
3	2429.40	2258.80	2266.80	0.249179
4	2246.30	2210.10	2218.90	0.153774
5	4069.00	4065.00	4073.00	0.148644
6	1538.30	1481.20	1489.20	0.283740
7	2319.50	2295.50	2303.50	0.131618
8	2197.80	2071.00	2079.00	0.181472
9	2456.90	2435.60	2443.60	0.182925
Mean	2399.46	2342.84	2350.84	0.228844

Seven Dials, London

Table D.9: Hypothesis 3 - Retweeted - Logistic regression results table - Seven Dials, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	702.55	699.86	707.86	0.317414
3	659.85	598.54	606.54	0.539462
4	1445.40	1429.00	1437.0	0.136393
5	1933.30	1932.70	1940.70	0.128079
6	851.03	846.65	854.65	0.311697
7	844.83	824.77	832.77	0.476072
8	1644.60	1638.10	1646.10	0.249393
9	720.65	685.15	693.15	0.398084
Mean	1100.28	1081.85	1089.85	0.319574

Oxford Street, London

Table D.10: Hypothesis 3 - Retweeted - Logistic regression results table - Oxford Street, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1215.30	1209.60	1217.60	0.279585
3	1366.90	1346.90	1354.90	0.102693
4	1423.30	1421.70	1429.70	0.128694
5	3279.30	3263.00	3271.00	0.131856
6	1232.20	1209.00	1217.00	0.267572
7	1137.30	1097.80	1105.80	0.230147
8	2548.20	2446.90	2454.90	0.562057
9	1129.40	1098.90	1106.90	0.482652
Mean	1722.74	1636.73	1644.73	0.273157

Covent Garden, London

Table D.11: Hypothesis 3 - Retweeted - Logistic regression results table - Covent Garden, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	2924.90	2808.30	2816.30	0.203876
3	2505.60	2264.10	2272.10	0.251757
4	6258.90	6159.90	6167.90	0.103572
5	2802.90	2770.00	2778.00	0.173915
6	1882.00	1746.50	1754.50	0.261529
7	4790.30	4630.40	4638.40	0.225827
8	1698.50	1633.00	1641.00	0.508780
9	2700.00	2281.70	2289.70	0.283066
Mean	3195.39	2974.24	3044.74	0.251540

London West End

Table D.12: Hypothesis 3 - Retweeted - Logistic regression results table - London West End.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	3220.90	2883.40	2891.40	0.510690
3	4854.50	4300.50	4308.50	0.215538
4	6653.50	5251.70	5259.70	0.305367
5	2692.10	2352.20	2360.20	0.285700
6	3395.50	2669.50	2677.50	0.303973
7	4979.00	4677.30	4685.30	0.155136
8	2342.90	2009.80	2017.80	0.262691
9	2667.60	2205.80	2213.80	0.262466
Mean	3850.75	3293.78	3301.78	0.286570

D.1.3 Retweet Count to Date

Regent Street, London

Table D.13: Hypothesis 3 - Retweet count to date - Logistic regression results table - Regent Street, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1713.40	1713.00	1721.00	0.372320
3	507.59	495.63	503.63	0.436336
4	1548.10	1542.60	1550.60	0.101453
5	2654.10	2637.90	2645.90	0.130494
6	1007.43	987.59	995.59	0.378057
7	1533.10	1532.00	1540.00	0.248034
8	1823.50	1806.80	1814.80	0.242936
9	1208.00	1149.20	1157.20	0.365130
Mean	1499.40	1483.09	1491.09	0.284345

Carnaby London

Table D.14: Hypothesis 3 - Retweet count to date - Logistic regression results table - Carnaby London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1938.50	1925.40	1933.40	0.499216
3	2429.40	2269.90	2277.90	0.245199
4	2246.30	2210.80	2218.80	0.153813
5	4069.00	4065.40	4073.40	0.148550
6	1538.30	1483.40	1491.40	0.282584
7	2319.50	2294.40	2302.40	0.132069
8	2197.80	2066.10	2074.10	0.183522
9	2456.90	2436.30	2444.30	0.182684
Mean	2399.46	2343.96	2351.96	0.228455

Seven Dials, London

Table D.15: Hypothesis 3 - Retweet count to date - Logistic regression results table - Seven Dials, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	702.55	699.37	707.37	0.317914
3	659.85	598.68	606.68	0.539342
4	1445.40	1442.10	1450.10	0.127990
5	1933.30	1928.60	1936.60	0.130160
6	851.03	845.13	853.13	0.313002
7	844.83	823.66	831.66	0.476812
8	1644.60	1637.40	1645.40	0.249787
9	720.65	685.32	693.22	0.397929
Mean	1100.28	1082.53	1090.52	0.319092

Oxford Street, London

Table D.16: Hypothesis 3 - Retweet count to date - Logistic regression results table - Oxford Street, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1215.30	1208.50	1216.50	0.280247
3	1366.90	1345.30	1353.30	0.103892
4	1423.30	1419.90	1427.90	0.129888
5	3729.30	3263.00	3271.00	0.131857
6	1232.20	1206.10	1214.10	0.269432
7	1137.30	1083.30	1091.30	0.240760
8	2548.20	2444.90	2452.90	0.562447
9	1129.40	1096.40	1104.40	0.483860
Mean	1722.74	1633.43	1641.43	0.275295

Covent Garden, London

Table D.17: Hypothesis 3 - Retweet count to date - Logistic regression results table - Covent Garden, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	2924.90	2808.40	2816.40	0.203835
3	2505.60	2265.10	2273.10	251423
4	6258.90	6159.40	6167.40	0.103656
5	2802.90	2770.50	2778.50	0.173750
6	1882.00	1758.10	1766.10	0.256437
7	4790.30	4633.40	4641.40	0.225281
8	1698.50	1632.10	1640.10	0.509070
9	2700.00	2283.90	2291.90	0.282340
Mean	3195.39	3038.86	3046.86	0.250729

London West End

Table D.18: Hypothesis 3 - Retweet count to date - Logistic regression results table - London West End.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	3220.90	2883.40	2891.40	0.510685
3	4854.50	4301.90	4309.90	0.215273
4	6653.50	5251.80	5259.80	0.305358
5	2692.10	2354.70	2362.70	0.284908
6	3395.50	2669.40	2677.40	0.303986
7	4979.00	4676.10	4684.10	0.155350
8	2342.90	2010.20	2018.20	0.262567
9	2667.60	2205.10	2213.10	0.262732
Mean	3850.75	3294.08	3302.08	0.286482

D.1.4 Retweeted to Date

Regent Street, London

Table D.19: Hypothesis 3 - Retweeted to date - Logistic regression results table - Regent Street, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1713.40	1712.40	1720.40	0.372582
3	507.59	495.63	503.63	0.426336
4	1548.10	1543.30	1551.30	0.100989
5	2654.10	2633.30	2641.30	0.132195
6	1007.43	986.71	994.71	0.378640
7	1533.10	1532.10	1540.10	0.247977
8	1823.50	1805.90	1813.90	0.243303
9	1208.00	1147.10	1155.10	0.366356
Mean	1499.40	1482.06	1490.06	0.284797

Carnaby London

Table D.20: Hypothesis 3 - Retweeted to date - Logistic regression results table - Carnaby London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1938.50	1924.80	1932.80	0.499392
3	2429.40	2264.20	2272.20	0.247241
4	2246.30	2210.80	2218.80	0.153805
5	4069.00	4066.90	4074.90	0.148187
6	1538.30	1483.20	1491.20	0.282917
7	2319.50	2296.10	2304.10	0.131387
8	2197.80	2069.50	2077.50	0.182158
9	2456.90	2437.80	2445.80	0.182158
Mean	2399.46	2344.16	2352.16	0.228372

Seven Dials, London

Table D.21: Hypothesis 3 - Retweeted to date - Logistic regression results table - Seven Dials, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	702.55	699.37	707.37	0.317914
3	659.85	597.86	605.86	0.540000
4	1445.40	1436.30	1444.30	0.131603
5	1933.30	1932.70	1940.70	0.128106
6	851.03	837.15	845.15	0.319832
7	844.83	823.66	831.66	0.476812
8	1644.60	1637.10	1645.10	0.249936
9	720.65	685.31	693.31	0.397937
Mean	1100.28	1081.18	1122.93	0.320268

Oxford Street, London

Table D.22: Hypothesis 3 - Retweeted to date - Logistic regression results table - Oxford Street, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	1215.30	1208.50	1216.50	0.280247
3	1366.90	1345.30	1353.30	0.103872
4	1423.30	1419.90	1427.90	0.129888
5	3279.30	3262.70	3270.70	0.131921
6	1232.20	1206.10	1214.10	0.269432
7	1137.30	1083.30	1091.30	0.240760
8	2548.20	2446.10	2454.10	0.562226
9	1129.40	1096.40	1104.40	0.483860
Mean	1722.74	1633.54	1641.54	0.275276

Covent Garden, London

Table D.23: Hypothesis 3 - Retweeted to date - Logistic regression results table - Covent Garden, London.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	2924.90	2808.40	2816.40	0.203854
3	2505.60	2265.30	2273.30	0.251332
4	6258.90	6161.00	6169.00	0.103394
5	2802.90	2770.50	2778.50	0.173760
6	1882.00	1756.90	1764.90	0.256960
7	4790.30	4632.10	4640.10	0.225511
8	1698.50	1632.50	1640.50	0.508919
9	2700.00	2282.50	2290.50	0.282796
Mean	3195.39	3038.65	3046.65	0.250816

London West End

Table D.24: Hypothesis 3 - Retweeted to date - Logistic regression results table - London West End.

Snapshot	Null Dev.	Residual Dev.	AIC	R^2
2	3220.90	2883.30	2891.30	0.501695
3	4854.50	4190.00	4198.00	0.236882
4	6653.50	5212.70	5220.70	0.310913
5	2692.10	2344.90	2352.90	0.287974
6	3395.50	2664.20	2672.20	0.305403
7	4979.00	4657.00	4665.00	0.159021
8	2342.90	2008.60	2016.60	0.263153
9	2667.60	2205.80	2213.80	0.262469
Mean	3850.75	3270.81	3278.81	0.290939

References

[1] Data Protection Act 1998, 1998. URL https://www.legislation.gov.uk/ukpga/1998/29.

[2] Data Protection Act 2018, 2018. URL https://www.legislation.gov.uk/ukpga/2018/12/.

[3] C. C. Aggarwal, A. Hinneburg, and D. A. Keim. On the Surprising Behavior of Distance Metrics in High Dimensional Space. In *International Conference on Database Theory*, pages 420–434, 2001. doi: 10.1007/3-540-44503-X{_}27.

[4] R. Agnihotri, R. Dingus, M. Y. Hu, and M. T. Krush. Social media: Influencing Customer Satisfaction in B2B Sales. *Industrial Marketing Management*, pages 172–180, 2015. doi: 10.1016/j.indmarman.2015.09.003.

[5] Alchemy API. Alchemy API, 2017. URL https://www.ibm.com/watson/alchemy-api.html.

[6] Y. Ardiansyah, P. Harrigan, G. N. Soutar, and T. M. Daly. Antecedents to Consumer Peer Communication through Social Advertising: A Self-Disclosure Theory Perspective. *Journal of Interactive Advertising*, 18(1):55–71, 2018. doi: 10.1080/15252019.2018.1437854.

[7] Association of Internet Researchers. Ethical Decision-Making and Internet Research. Recommendations from the AOiR Ethics Working Committee. Technical report, Association of Internet Researchers, 2002. URL http://www.aoir.org/reports/ethics.pdf.

[8] Association of Internet Researchers. Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0). Technical report, Association of Internet Researchers, 2012. URL http://www.aoir.org/reports/ethics2.pdf.

[9] Association of Internet Researchers. Internet Research : Ethical Guidelines 3.0 Association of Internet Researchers. Technical report, 2019. URL http://www.aoir.org/reports/ethics3.pdf.

[10] O. Atkins. Boomers go digital: how consumer behaviour has changed in 2020, 12 2020. URL https://www.thedrum.com/news/2020/12/07/boomers-go-digital-how-consumer-behaviour-has-changed-2020.

[11] A. Backiel, Y. Verbinnen, B. Baesens, and G. Claeskens. Combining Local and Social Network Classifiers to Improve Churn Prediction. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 651–658, 2015. doi: 10.1145/2808797.2808850.

[12] M. Baker. *Market Strategy and Management*. Palgrave Macmillan, London, England, 3rd edition, 2000.

[13] A. Banerjee, U. B. Chitnis, S. L. Jadhav, J. S. Bhawalkar, and S. Chaudhury. Hypothesis Testing, Type I and Type II Errors. *Industrial Psychiatry Journal*, 18(2):127–131, 2009. doi: 10.4103/0972-6748.62274.

[14] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *Third International AAAI Conference on Weblogs and Social Media*, pages 361–362, 2009. doi: 10.1136/qshc.2004.010033.

[15] A. Bavelas. A Mathematical Model for Group Structures. *Human Organization*, 7(3):16–30, 1948. doi: 10.17730/humo.7.3.f4033344851gl053.

[16] A. Bavelas. Communication Patterns in Task-Oriented Groups. *The Journal of the Acoustical Society of America*, 22(6):725–730, 1950. doi: 10.1121/1.1906679.

[17] C. H. Bélanger, S. Bali, and B. Longden. How Canadian Universities Use Social Media to Brand Themselves. *Tertiary Education and Management*, 20(1):14–29, 2014. doi: 10.1080/13583883.2013.852237.

[18] R. Benbunan-Fich and S. Altschuller. Web Presence Transformations in the 1990s: An Analysis of Press Releases. *IEEE Transactions on Professional Communication*, 48(2):131–146, 2005. doi: 10.1109/TPC.2005.849647.

[19] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), 2008. doi: 10.1088/1742-5468/2008/10/P10008.

[20] Bluewater. Bluewater Shopping and Leisure Destination, 2019. URL `https://bluewater.co.uk`.

[21] H. Bonner. *Group Dynamics: Principles and Applications*. Ronald Press, New York, New York, USA, 1959.

[22] Boxpark Shoreditch. Boxpark Shoreditch, 2018. URL `https://www.boxpark.co.uk/shoreditch`.

[23] R. E. Boyatzis. *Transforming Qualitative Information: Thematic Analysis and Code Development*. Sage Publications Ltd, London, 1998.

[24] d. boyd. Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life. In *Youth, Identity, and Digital Media*. MIT Press, 2007.

[25] d. boyd and K. Crawford. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15(5):662–679, 2012. doi: 10.1080/1369118x.2012.678878.

[26] d. boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, pages 1–10, 2010. doi: 10.1109/HICSS.2010.412.

[27] V. Braun and V. Clarke. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006. doi: 10.1191/1478088706qp063oa.

[28] V. Braun and V. Clarke. *Successful Qualitative Research. A Practical Guide for Beginners*. Sage, Los Angeles, 2013.

[29] V. Braun and V. Clarke. Reflecting on Reflexive Thematic Analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4):588–597, 2019. doi: 10.1080/2159676X.2019.1628806.

[30] V. Braun, V. Clarke, N. Hayfield, and G. Terry. Thematic Analysis. In P. Liamputtong, editor, *Handbook of Research Methods in Health Social Sciences*. Springer Nature, 2018. doi: 10.13140/RG.2.1.3418.5682.

[31] Brent Cross. Brent Cross Shopping Centre London, 2018. URL `https://www.brentcross.co.uk`.

[32] British Psychological Society. Ethics Guidelines for Internet-Mediated Research. Technical report, British Psychological Society, 2013.

[33] British Psychological Society. Ethics Guidelines for Internet-mediated Research. Technical report, British Psychological Society, 2017.

[34] British Sociological Association. Statement of Ethical Practice. Technical report, The British Sociological Association, 2017.

[35] British Sociological Association. Researching Online Forums Online Forums - Ethics Case Study 1. Technical report, The British Sociological Association, 2017.

[36] British Sociological Association. New Methodological Approaches to Research Using Twitter - Ethics Case Study 4. Technical report, The British Sociological Association, 2017.

[37] British Sociological Association. Researching Social Machines Social Machines - Ethics Case Study 3. Technical report, The British Sociological Association, 2017.

[38] British Sociological Association. Ethics Guidelines and Collated Resources for Digital Research. Technical report, British Sociological Association, 2017.

[39] British Sociological Association. Using Twitter for Criminology Research - Ethics Case Study 2. Technical report, British Sociological Association, 2017.

[40] British Sociological Association. Mixed Methods Study Involving Young People - Ethics Case Study 7. Technical report, British Sociological Association, 2017.

[41] British Sociological Association. Open Data and Democratic Governance: Policies, Platforms and Practices - Ethics Case Study 6. Technical report, British Sociological Assocation, 2017.

[42] A. Bruckman. Opportunities and Challenges in Methodology and Ethics. In M. D. Johns, S. L. S. Chen, and G. J. Hall, editors, *Online Social Research: Methods, Issues and Ethics*. Peter Lang, New York, New York, USA, 2004.

[43] Bullring & Grand Central. Bullring & Grand Central, 2018. URL `https://www.bullring.co.uk`.

[44] L. Burbach, P. Halbach, M. Ziefle, and A. Calero Valdez. Who Shares Fake News in Online Social Networks? In *UMAP '19: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, pages 234–242, 2019. doi: 10.1145/3320435.3320456.

[45] P. Burnap, R. K. Gibson, L. Sloan, R. Southern, and M. L. Williams. 140 Characters to Victory?: Using Twitter to Predict the UK 2015 General Election. *SSRN Electronic Journal*, 2015. doi: 10.2139/ssrn.2603433.

[46] J. Canning and P. A. Found. The Effect of Resistance in Organizational Change Programmes: A Study of a Lean Transformation. *International Journal of Quality and Service Sciences*, 7 (2):274–295, 2015. doi: 10.1108/IJQSS-02-2015-0018.

[47] D. Cartwright and A. Zander. *Group Dynamics: Research and Theory*. Harper & Row, New York, New York, USA, 1968.

[48] J. Colliander, M. Dahlen, and E. Modig. Twitter for Two: Investigating the Effects of Dialogue with Customers in Social Media. *International Journal of Advertising*, 34(2):181–194, 2015. doi: 10.1080/02650487.2014.996197.

[49] J. W. Creswell and J. D. Creswell. *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, 2017.

[50] N. Crossley, E. Bellotti, G. Edwards, M. G. Everett, J. Koskinen, and M. Tranmer. *Social Network Analysis for Ego-Nets*. SAGE Publications Ltd, London, 2015. doi: 10.4135/9781473911871.

[51] G. Csardi and T. Nepusz. The IGraph Software Package for Complex Network Research. *InterJournal, Complex Systems*, 1695(5):1–9, 2006.

[52] A. D'Arcy and T. M. Young. Ethics and social media: Implications for sociolinguistics in the networked public. *Journal of Sociolinguistics*, 16:534–546, 2012.

[53] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi. Social Ties and their Relevance to Churn in Mobile Telecom Networks. In *Proceedings of the 11th International Conference on Extending Database Technology*, pages 668–677, 2008. doi: 10.1145/1353343.1353424.

[54] M. De Choudhury, H. Sundaram, A. John, D. D. Seligmann, and A. Kelliher. "Birds of a Feather": Does User Homophily Impact Information Diffusion in Social Media? *arXiv preprint*, 2010.

[55] C. Dijkmans, P. Kerkhof, A. Buyukcan-Tetik, and C. J. Beukeboom. Online Conversation and Corporate Reputation: A Two-Wave Longitudinal Study on the Effects of Exposure to the Social Media Activities of a Highly Interactive Company. *Journal of Computer-Mediated Communication*, 20(6):632–648, 2015. doi: 10.1111/jcc4.12132.

[56] D. DiNucci. Fragmented Future. *Print*, 53(4):32–33, 1999.

[57] R. Dolan, J. Conduit, C. Frethey-Bentham, J. Fahy, and S. Goodman. Social media engagement behavior. *European Journal of Marketing*, 53(10), 10 2019. ISSN 0309-0566. doi: 10.1108/EJM-03-2017-0182.

[58] U. Droftina, M. Štular, and A. Košir. A Diffusion Model for Churn Prediction Based on Sociometric Theory. *Advances in Data Analysis and Classification*, 9(3):341–365, 2015. doi: 10.1007/s11634-014-0188-0.

[59] E. Durkheim and G. E. G. Caitlin (Translator). *The Rules of Sociological Method*. The Free Press, New York, New York, USA, 8th edition, 1938.

[60] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.

[61] N. Edelmann. Reviewing the Definitions of "Lurkers" and Some Implications for Online Research. *Cyberpsychology, Behavior, and Social Networking*, 16(9), 9 2013. doi: 10.1089/cyber.2012.0362.

[62] F. Eskandanian, N. Sonboli, and B. Mobasher. Power of the few: Analyzing the impact of influential users in collaborative recommender systems. *ACM UMAP 2019 - Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, pages 225–233, 2019. doi: 10.1145/3320435.3320464.

[63] European Parliament and Council of European Union. Regulation (EU) 2016/679, 2016. URL `https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN`.

[64] Facebook. Facebook for Developers, 2018. URL `https://developers.facebook.com`.

[65] J. A. M. Farver. Aggressive Behavior in Preschoolers' Social Networks: Do Birds of a Feather Flock Together? *Early Childhood Research Quarterly*, 11(3):333–350, 1996. doi: 10.1016/S0885-2006(96)90011-3.

[66] R. Ferraro, J. R. Bettman, and T. L. Chartrand. The Power of Strangers: The Effect of Incidental Consumer Brand Encounters on Brand Choice. *Journal of Consumer Research*, 35(5): 729–741, 2009. doi: 10.1086/592944.

[67] C. Fiesler and N. Proferes. "Participant" Perceptions of Twitter Research Ethics. *Social Media and Society*, 4(1), 2018. doi: 10.1177/2056305118763366.

[68] R. E. Frank, W. F. Massy, and Y. Wind. *Market Segmentation*. Prentice Hall, New Jersey, USA, 1972.

[69] L. C. Freeman. The Sociological Concept of 'Group': An Empirical Test of Two Models. *American Journal of Sociology*, 98:152–166, 1992.

[70] L. C. Freeman. *The Development of Social Network Analysis. A Study in the Sociology of Science*. Empirical Press, Vancouver, Canada, 2004.

[71] N. E. Friedkin. *A Structural Theory of Social Influence*. Cambridge University Press, 9 1998. doi: 10.1017/CBO9780511527524.

[72] N. E. Friedkin and E. C. Johnsen. Social Positions in Influence Networks. *Social Networks*, 19(3):209–222, 1997. doi: 10.1016/S0378-8733(96)00298-5.

[73] N. E. Friedkin and E. C. Johnsen. *Social Influence Network Theory*. Cambridge University Press, Cambridge, 2011. doi: 10.1017/CBO9780511976735.

[74] C. Fuchs. 'Dear Mr. Neo-Nazi, Can You Please Give Me Your Informed Consent so That I Can Quote Your Fascist Tweet?': Questions of Social Media Research Ethics in Online Ideology Critique. In G. Meikle, editor, *The Routledge Companion To Media And Activism*, pages 385–394. Routledge, Abingdon, United Kingdom, 2018. doi: 10.4324/9781315475059-41.

[75] Q. Gao and C. Feng. Branding with Social Media: User Gratifications, Usage Patterns, and Brand Message Content Strategies. *Computers in Human Behavior*, 63:868–890, 2016. doi: 10.1016/j.chb.2016.06.022.

[76] M. Gasparini, R. Clarisó, M. Brambilla, and J. Cabot. Participation Inequality and the 90-9-1 Principle in Open Source. In *Proceedings of the 16th International Symposium on Open Collaboration*, New York, NY, USA, 8 2020. ACM. ISBN 9781450387798. doi: 10.1145/3412569.3412582.

[77] D. Gayo-Avello. Social Media, Democracy, and Democratization. *IEEE Multimedia*, 22(2): 10–16, 2015. doi: 10.1109/MMUL.2015.47.

[78] M. Girvan and M. E. J. Newman. Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002. doi: 10.1073/pnas.122653799.

[79] N. Glady, B. Baesens, and C. Croux. Modeling Churn using Customer Lifetime Value. *European Journal of Operational Research*, 197(1):402–411, 2009. doi: 10.1016/j.ejor.2008.06.027.

[80] B. Gleason and S. von Gillern. Digital Citizenship with Social Media: Participatory Practices of Teaching and Learning in Secondary Education. *Journal of Educational Technology & Society*, 21(1):200–212, 2018.

[81] Greg Satell. A Look Back At Why Blockbuster Really Failed And Why It Didn't Have To, 2015. URL `http://www.forbes.com/sites/gregsatell/2014/09/05/a-look-back-at-why-blockbuster-really-failed-and-why-it-didnt-have-to`.

[82] M. Guillén, J. P. Nielsen, T. H. Scheike, and A. M. Pérez-Marín. Time-Varying Effects in the Analysis of Customer Loyalty: A Case Study in Insurance. *Expert Systems with Applications*, 39(3):3551–3558, 2012. doi: 10.1016/j.eswa.2011.09.045.

[83] C. C. Günther, I. F. Tvete, K. Aas, G. I. Sandnes, and O. Borgan. Modelling and Predicting Customer Churn from an Insurance Company. *Scandinavian Actuarial Journal*, 1(1):58–71, 2014. doi: 10.1080/03461238.2011.636502.

[84] S. Haefliger, E. Monteiro, D. Foray, and G. von Krogh. Social Software and Strategy. *Long Range Planning*, 44(5-6):297–316, 2011. doi: 10.1016/j.lrp.2011.08.001.

[85] A. M. Harris, D. Gómez-Zará, L. A. Dechurch, and N. S. Contractor. Joining together online: The trajectory of CSCW scholarship on group formation. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 2019. ISSN 25730142. doi: 10.1145/3359250.

[86] G. Harris. *The Hottest Ideas in Word of Mouth Advertising. What Works, What Doesn't, What's Fake and Why*. Management Books, Essex, United Kingdom, 2006.

[87] W. He, F. K. Wang, Y. Chen, and S. Zha. An Exploratory Investigation of Social Media Adoption by Small Businesses. *Information Technology and Management*, 18(2):149–160, 2017. doi: 10.1007/s10799-015-0243-3.

[88] I. Himelboim, S. Mccreery, and M. Smith. Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter. *Journal of Computer-Mediated Communication*, 18(2):40–60, 2013. doi: 10.1111/jcc4.12001.

[89] D. L. Hoffman and T. P. Novak. Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations. *Journal of Marketing*, 60(3):50–68, 1996. doi: 10.2307/1251841.

[90] D. L. Hoffman, T. P. Novak, and P. Chatterjee. Commercial Scenarios for the Web: Opportunities and Challenges. *Journal of Computer-Mediated Communication*, 1(3), 1995. doi: 10.1111/j.1083-6101.1995.tb00165.x.

[91] M. A. Hogg. *The Social Psychology of Group Cohesiveness: From Attraction to Social Identity*. New York University Press, New York, New York, USA, 1992.

[92] G. C. Homans. *The Human Group*. Harcourt, Brace and World, New York, New York, USA, 1950.

[93] G. Hooley, N. F. Piercy, and B. Nicoulaud. *Marketing Strategy & Competitive Marketing*. Prentice Hall, London, United Kingdom, 5th edition, 2011.

[94] M. Horsnell. How to utilise IGTV to grow your brand, 2 2021. URL `https://www.thedrum.com/opinion/2021/02/03/how-utilise-igtv-grow-your-brand`.

[95] T. Hoskins. Parametric and Nonparametric: Demistifying the terms. Technical report, Mayo Clinic, 2012.

[96] E. R. Hubbell. Leading Through Evidence and Experience: How Social Media Provides a Real-Time Platform for Disseminating What Works in Education. *Australian Educational Leader*, 40(3):16–18, 2018.

[97] J. M. Hudson and A. Bruckman. ”Go away”: Participant objections to being studied and the ethics of chatroom research. *Information Society*, 20(2):127–139, 4 2004. doi: 10.1080/01972240490423030.

[98] S. Hudson, L. Huang, M. S. Roth, and T. J. Madden. The Influence of Social Media Interactions on Consumer-Brand Relationships: A Three-Country Study of Brand Perceptions and Marketing Behaviors. *International Journal of Research in Marketing*, 33(1):27–41, 2016. doi: 10.1016/j.ijresmar.2015.06.004.

[99] Intu. intu Metrocentre, Gateshead, 2019. URL `https://intu.co.uk/metrocentre`.

[100] A. K. Jain. Data Clustering: 50 Years Beyond K-Means. *Pattern Recognition Letters*, 31(8): 651–666, 2010. doi: 10.1016/j.patrec.2009.09.011.

[101] M. R. Jalilvand, S. S. Esfahani, and N. Samiei. Electronic Word-of-Mouth: Challenges and Opportunities. *Procedia Computer Science*, 3:42–46, 2011. doi: 10.1016/j.procs.2010.12.008.

[102] J. J. P. Jansen, F. A. J. Van Den Bosch, and H. W. Volberda. Managing Potential and Realized Absorptive Capacity: How Do Organizational Antecedents Matter? *Academy of Management Journal*, 48(6):999–1015, 2005. doi: 10.5465/AMJ.2005.19573106.

[103] D. B. Kandel. Homophily, Selection, and Socialization in Adolescent Friendships. *Journal of Sociology*, 84(2):427–436, 1978.

[104] A. A. Khan, S. Jamwal, and M. M. Sepehri. Applying Data Mining to Customer Churn Prediction in an Internet Service Provider. *International Journal of Computer Applications*, 9(7): 8–14, 2010. doi: 10.5120/1400-1889.

[105] M. L. Khan. Social Media Engagement: What Motivates User Participation and Consumption on YouTube? *Computers in Human Behavior*, 66:236–247, 2017. doi: 10.1016/j.chb.2016.09.024.

[106] A. J. Kim and E. Ko. Impacts of Luxury Fashion Brand's Social Media Marketing on Customer Relationship and Purchase Intention. *Journal of Global Fashion Marketing*, 1(3):164–171, 2010. doi: 10.1080/20932685.2010.10593068.

[107] D. Knoke and S. Yang. *Social Network Analysis*. SAGE Publications, London, United Kingdom, 2nd edition, 2008.

[108] Kollock. Peter and M. Smith. Managing the virtual commons. In S. Herring, editor, *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives*, pages 109–128. John Benjamins, Amsterdam, 1996.

[109] P. Kotler. *Marketing Management: Analysis, Planning, Implementation, and Control*. Prentice Hall, 1988.

[110] P. Kotler and K. L. Keller. *Marketing Management*. Pearson Education Limited, Harlow, England, 15th edition, 2016.

[111] Kotter Interntaional. Barriers to Change: The Real Reason Behind the Kodak Downfall, 2012. URL `http://www.forbes.com/sites/johnkotter/2012/05/02/barriers-to-change-the-real-reason-behind-the-kodak-downfall`.

[112] N. Krüger, S. Stieglitz, and T. Potthoff. Brand Communication In Twitter - A Case Study On Adidas. In *Proceedings of the 2012 Pacific Asia Conference on Information Systems*, 2012.

[113] E. Lahuerta-Otero and R. Cordero-Gutiérrez. Looking for the Perfect Tweet. The Use of Data Mining Techniques to Find Influencers on Twitter. *Computers in Human Behavior*, 64:575–583, 2016. doi: 10.1016/j.chb.2016.07.035.

[114] D. Langdridge and G. Hagger-Johnson. *Introduction to Research Methods and Data Analysis in Psychology*. Pearson Prentice Hall, 2nd edition, 2009.

[115] B. Larivière and D. Van Den Poel. Investigating the Role of Product Features in Preventing Customer Churn, by Using Survival Analysis and Choice Modeling: The Case of Financial Services. *Expert Systems with Applications*, 27(2):277–285, 2004. doi: 10.1016/j.eswa.2004.02.002.

[116] K. Lewin. *Field Theory in Social Science*. Harper, New York, New York, USA, 1951.

[117] Y. Li and Y. Xie. Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement. *Journal of Marketing Research*, 57(1), 2 2020. ISSN 0022-2437. doi: 10.1177/0022243719881113.

[118] S. Lindgren. *Data Theory. Interpretive Sociology and Computational Methods*. Polity Press, Cambridge, United Kingdom, 2020.

[119] A. Lipsman, G. Mudd, M. Rich, and S. Bruich. The Power of "Like". *Journal of Advertising Research*, 52(1):40–52, 2012. doi: 10.2501/JAR-52-1-040-052.

[120] X. Liu, A. C. Burns, and Y. Hou. An Investigation of Brand-Related User-Generated Content on Twitter. *Journal of Advertising*, 46(2):236–247, 2017. doi: 10.1080/00913367.2017.1297273.

[121] Y. Liu, J. Venkatanathan, J. Goncalves, E. Karapanos, and V. Kostakos. Modeling What Friendship Patterns on Facebook Reveal About Personality and Social Capital. *ACM Transactions on Computer-Human Interaction*, 21(3):1–20, 2014. doi: 10.1145/2617572.

[122] LNER. London North Eastern Railway Booking Site, 2019. URL `http://www.facebook.com/LNERail`.

[123] E. Locatelli. Ethics of Social Media Research: State of the Debate and Future Challenges. *Second International Handbook of Internet Research*, pages 1–22, 2018. doi: 10.1007/978-94-024-1202-4{_}25-1.

[124] London North Eastern Railway. Twitter - London North Eastern Railyway, 2019. URL `https://twitter.com/LNER`.

[125] D. L. Loudon and A. J. Della Biita. *Consumer Behavior: Concepts and Applications*. McGraw-Hill, 1984.

[126] S. a. Macskassy and M. Michelson. Why Do People Retweet? Anti-Homophily Wins the Day! In *Proceedings of the Fifth International Conference on Weblogs and Social Media*, pages 209–216, 2011.

[127] J. Mahoney, S. Lawson, and R. Stone. "What do you Think of the Return of Dungarees?". In *Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, pages 1927–1932, New York, New York, USA, 2014. ACM Press. doi: 10.1145/2559206.2581299.

[128] A. E. Marwick and d. boyd. I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media and Society*, 13(1):114–133, 2011. doi: 10.1177/1461444810365313.

[129] L. McAlister, G. Sonnier, and T. Shively. The Relationship Between Online Chatter and Firm Value. *Marketing Letters*, 23(1):1–12, 2012. doi: 10.1007/s11002-011-9142-5.

[130] J. McCarthy, J. Rowley, C. J. Ashworth, and E. Pioch. Managing Brand Presence Through Social Media: The Case of UK Football Clubs. *Internet Research*, 24(2):181–204, 2014. doi: 10.1108/IntR-08-2012-0154.

[131] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):415–444, 2002. doi: 10.1146/annurev.soc.27.1.415.

[132] N. Mehta and A. Atreja. Online Social Support Networks. In M. S. Chisolm, editor, *Social Media in Medicine*, pages 36–41. Routledge, 2018. doi: 10.4324/9781315619941-5.

[133] J. Metcalf and K. Crawford. Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society*, 3(1):205395171665021, 1 2016. doi: 10.1177/2053951716650211.

[134] N. Michaelidou, N. T. Siamagka, and G. Christodoulides. Usage, Barriers and Measurement of Social Media Marketing: An Exploratory Investigation of Small and Medium B2B Brands. *Industrial Marketing Management*, 40(7):1153–1159, 2011. doi: 10.1016/j.indmarman.2011.09.009.

[135] S. Monika, C. Anand, and R. K. Gnanamurthy. Analyzing the User Profile Linkage across Different Social Network Platforms. *International Journal of Computer Science and Engineering Communications*, 4(2):1378–1383, 2016.

[136] J. L. Moreno. *Who Shall Survive: A New Approach to the Problem of Human Interrelations*. Nervous and Mental Disease Publshing Company, Washington D.C, USA, 1934.

[137] J. L. Moreno. *Sociometry, Experimental Method and the Science of Society: An Approach to New Political Orientation*. Beacon House, New York, New York, USA, 1951.

[138] J. L. Moreno. *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. Bacon House, New York, New York, USA, 1953.

[139] J. L. Moreno and H. H. Jennings. Statistics of Social Configurations. *Sociometry*, 1(3):342–374, 1938.

[140] M. A. Moreno, N. Goniu, P. S. Moreno, and D. Diekema. Ethics of Social Media Research: Common Concerns and Practical Considerations. *Cyberpsychology, Behavior, and Social Networking*, 16(9):708–713, 9 2013. doi: 10.1089/cyber.2012.0334.

[141] M. Morris and C. Ogan. The Internet as Mass Medium. *Journal of Computer-Mediated Communication*, 1(4), 6 1996. ISSN 10836101. doi: 10.1111/j.1083-6101.1996.tb00174.x.

[142] M. Naaman, J. Boase, and C.-H. Lai. Is it Really About Me? Message Content in Social Awareness Streams. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, 2010. doi: 10.1371/journal.pcbi.1003109.

[143] M. E. J. Newman. *Networks. An Introduction*. Oxford University Press, Oxford, United Kingdom, 2010.

[144] M. E. J. Newman and M. Girvan. Finding and Evaluating Community Structure in Networks. *Physical Review E*, 69(2), 2004. doi: 10.1103/PhysRevE.69.026113.

[145] N. Newman. Journalism, Media and Technology Predictions 2016. Technical report, Reuters Institute, 2016.

[146] J. Nielsen. Participation Inequality: The 90-9-1 Rule for Social Features, 2006. URL `https://www.nngroup.com/articles/participation-inequality/`.

[147] H. Nissenbaum. Protecting Privacy in an Information Age: The Problem of Privacy in Public. *Law and Philosophy*, 17(5):559–596, 1998.

[148] H. Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, 79, 2004. URL `https://heinonline.org/HOL/Page?handle=hein.journals/washlr79&id=129&div=16&collection=journals`.

[149] B. Nonnecke, J. Preece, and D. Andrews. What lurkers and posters think of each other. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*. IEEE, 2004. doi: 10.1109/HICSS.2004.1265462.

[150] B. Nonnecke, D. Andrews, and J. Preece. Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, 6(1), 1 2006. doi: 10.1007/s10660-006-5985-x.

[151] A. Olteanu, C. Castillo, F. Diaz, and E. Kıcıman. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, 2(13):13, 7 2019. doi: 10.3389/fdata.2019.00013.

[152] T. O'Reilly. Web 2.0: Compact Definition, 2005. URL `http://radar.oreilly.com/2005/10/web-20-compact-definition.html`.

[153] M. Oskarsdottir, C. Bravo, W. Verbeke, C. Sarraute, B. Baesens, and J. Vanthienen. A Comparative Study of Social Network Classifiers for Predicting Churn in the Telecommunication Industry. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1151–1158, 2016. doi: 10.1109/ASONAM.2016.7752384.

[154] Oxford Street. Oxford Street - Shop London, 2018. URL `https://www.oxfordstreet.co.uk`.

[155] A. Palmer. *Principles of Marketing*. Oxford University Press, Oxford, United Kingdom, 2000.

[156] W. Pan. Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics*, 57 (1):120–125, 2001. doi: 10.1111/j.0006-341X.2001.00120.x.

[157] E. Pancer, V. Chandler, M. Poole, and T. J. Noseworthy. How Readability Shapes Social Media Engagement. *Journal of Consumer Psychology*, 29(2), 4 2019. ISSN 1057-7408. doi: 10.1002/jcpy.1073.

[158] C. Peersman, W. Daelemans, and L. Van Vaerenbergh. Predicting Age and Gender in Online Social Networks. In *Proceedings of the 3rd International Workshop on Search and Mining User-Generated Contents*, 2011. doi: 10.1145/2065023.2065035.

[159] M. Pennacchiotti and A.-M. Popescu. A Machine Learning Approach to Twitter User Classification. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 281–288, 2011.

[160] Pew Research Center. Demographics of Social Media Users in 2016, 2016. URL `http://www.pewinternet.org/2016/11/11/social-media-update-2016`.

[161] Pew Research Center. Social Media Use 2018: Demographics and Statistics, 2018. URL `http://www.pewinternet.org/2018/03/01/social-media-use-in-2018`.

[162] J. Phua, S. V. Jin, and J. J. Kim. Gratifications of Using Facebook, Twitter, Instagram, or Snapchat to Follow Brands: The Moderating Effect of Social Comparison, Trust, Tie Strength, and Network Homophily on Brand Identification, Brand Engagement, Brand Commitment, and Membership Intentio. *Telematics and Informatics*, 34(1):412–424, 2017. doi: 10.1016/j.tele.2016.06.004.

[163] R Core Team. R: A Language and Environment for Statistical Computing, 2017. URL `https://www.r-project.org`.

[164] D. Radosavljevik, P. V. D. Putten, and K. K. Larsen. The Impact of Experimental Setup in Prepaid Churn Prediction for Mobile Telecommunications: What to Predict, for Whom and Does the Customer Experience Matter? *Transactions on Machine Learning and Data Mining*, 3(2):80–99, 2010.

[165] D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta. Classifying Latent User Attributes in Twitter. In *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents*, pages 37–44, New York, New York, USA, 2010. ACM Press. doi: 10.1145/1871985.1871993.

[166] R. Rentschler. Museum Marketing: Understanding Different Types of Audiences. In F. Kerrigan, P. Fraser, and M. Ozbilgin, editors, *Arts Marketing*, chapter 8, pages 149–168. Elsevier Butterworth-Heinemann, Oxford, United Kingdom, 2004.

[167] Y. Richter, E. Yom-Tov, and N. Slonim. Predicting Customer Churn in Mobile Networks through Analysis of Social Groups. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pages 732–741, 2010. doi: 10.1137/1.9781611972801.64.

[168] M. Saravanan and G. S. Vijay Raajaa. A Graph-Based Churn Prediction Model for Mobile Telecom Networks. *Lecture Notes in Computer Science*, pages 367–382, 2012. doi: 10.1007/978-3-642-35527-1{_}31.

[169] D. Schlagwein and M. Hu. How and Why Organisations Use Social Media: Five Use Types and their Relation to Absorptive Capacity. *Journal of Information Technology*, 32(2):194–209, 2017. doi: 10.1057/jit.2016.7.

[170] H. Shahbaznezhad, R. Dolan, and M. Rashidirad. The Role of Social Media Content Format and Platform in Users' Engagement Behavior. *Journal of Interactive Marketing*, 53, 2 2021. ISSN 10949968. doi: 10.1016/j.intmar.2020.05.001.

[171] A. Sharma and D. Cosley. Studying and Modeling the Connection between People's Preferences and Content Sharing. In *Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing*, pages 1246–1257, 2015. doi: 10.1145/2675133.2675151.

[172] S. Sharma and H. V. Verma. Social Media Marketing: Evolution and Change. In *Social Media Marketing*, pages 19–36. Springer Singapore, Singapore, 2018. doi: 10.1007/978-981-10-5323-8{_}2.

[173] M. E. Shaw. *Group Dynamics: The Psychology of Small Group Behavior*. McGraw-Hill, New York, New York, USA, 1976.

[174] J. N. Sheth. How Social Media Will Impact Marketing Media. In G. Heggde and G. Shainesh, editors, *Social Media Marketing*, pages 3–18. Palgrave Macmillan, Singapore, 2018. doi: 10.1007/978-981-10-5323-8{_}1.

[175] G. Simmel. *The Sociology of Georg Simmel*. Collier Macmillan Publishers, London, United Kingdom, 1950.

[176] L. Sloan, J. Morgan, W. Housley, M. Williams, A. Edwards, P. Burnap, and O. Rana. Knowing the Tweeters: Deriving Sociologically Relevant Demographics from Twitter. *Sociological Research Online*, 18(3):1–11, 2013. doi: 10.5153/sro.3001.

[177] L. Sloan, J. Morgan, P. Burnap, and M. Williams. Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data. *PLoS ONE*, 10(3):1–20, 2015. doi: 10.1371/journal.pone.0115545.

[178] J. Smith, M. Larkin, and P. Flowers. *Interpretative Phenomenological Analysis: Theory, Method and Research*. SAGE Publications, 2009.

[179] R. A. Smith and E. L. Fink. Understanding the Influential People and Social Structures Shaping Compliance. *Journal of Social Structure*, 16:1–15, 2015.

[180] W. R. Smith. Product Differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of Marketing*, 21(1):3–8, 1956. doi: 10.2307/1247695.

[181] C. L. Snelson. Qualitative and mixed methods social media research: A review of the literature. *International Journal of Qualitative Methods*, 15(1):1–15, 2 2016. doi: 10.1177/1609406915624574.

[182] A. T. Stephen and O. Toubia. Deriving Value from Social Commerce Networks. *Journal of Marketing Research*, 47(2):215–228, 2010. doi: 10.1509/jmkr.47.2.215.

[183] R. M. Stogdill. *Individual Behavior and Group Achievement: A Theory, the Experimental Evidence*. Oxford University Press, New York, New York, USA, 1959.

[184] A. L. Strauss and J. M. Corbin. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications Ltd, 1998.

[185] E. Sun, I. Rosenn, C. A. Marlow, and T. M. Lento. Gesundheit! Modeling Contagion through Facebook News Feed Mechanics of Facebook Page Diffusion. In *Proceedings of the 3rd International ICWSM Conference*, number 2000, pages 146–153, 2009.

[186] N. Sun, P. P.-L. Rau, and L. Ma. Understanding lurkers in online communities: A literature review. *Computers in Human Behavior*, 38, 9 2014. ISSN 07475632. doi: 10.1016/j.chb.2014.05.022.

[187] T. Suomela, F. Chee, B. Berendt, and G. Rockwell. Applying an Ethics of Care to Internet Research: Gamergate and Digital Humanities. *Digital Studies/Le champ numérique*, 9(1), 2 2019. doi: 10.16995/dscn.302.

[188] Text Razor. Text Razor - The Natural Language Processing API, 2017. URL `https://www.textrazor.com`.

[189] R. Tibshirani, G. Walther, and T. Hastie. Estimating the Number of Clusters in a Data Set via the gap Statistic. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 63 (2):411–423, 2001. doi: 10.1111/1467-9868.00293.

[190] G. Torkzadeh, J. C. J. Chang, and G. W. Hansen. Identifying Issues in Customer Relationship Management at Merck-Medco. *Decision Support Systems*, 42(2):1116–1130, 2006. doi: 10.1016/j.dss.2005.10.003.

[191] L. Townsend and C. Wallace. Social Media Research: A Guide to Ethics. Technical report, The University of Aberdeen, 2016.

[192] Trinity Leeds. Trinity Leeds - 120 Shops, Restaurants, Bars and Cinema, 2018. URL `https://trinityleeds.com`.

[193] M. Trusov, R. E. Bucklin, and K. Pauwels. Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site. *Journal of Marketing*, 73(5):90–102, 2009. doi: 10.1509/jmkg.73.5.90.

[194] Twitter. Twitter - API Reference Index, 2017. URL `https://developer.twitter.com/en/docs/api-reference-index`.

[195] T. van Mierlo. The 1% Rule in Four Digital Health Social Networks: An Observational Study. *Journal of Medical Internet Research*, 16(2), 2 2014. ISSN 14388871. doi: 10.2196/jmir.2966.

[196] W. Verbeke, D. Martens, and B. Baesens. Social Network Analysis for Customer Churn Prediction. *Applied Soft Computing Journal*, 14:431–446, 2014. doi: 10.1016/j.asoc.2013.09.017.

[197] J. Vitak, K. Shilton, and Z. Ashktorab. Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community. In *Computer-Supported Cooperative Work and Social Computing 2016*, 2016. doi: 10.1145/2818048.2820078.

[198] C. Wagner and A. Majchrzak. Enabling Customer-Centricity using Wikis and the Wiki Way. *Journal of Management Information Systems*, 23(3):17–43, 2006. doi: 10.2753/MIS0742-1222230302.

[199] N. Walliman. *Social Research Methods*. SAGE Publications Ltd, London, United Kingdom, 2nd edition, 2017.

[200] S. Wasserman and K. Faust. *Social Network Analysis. Methods and Applications*. Cambridge University Press, Cambridge, United Kingdom, 1994.

[201] We Are Social UK. Digital 2019: Global Internet Use Accelerates, 2019. URL `https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates`.

[202] M. Wedel and W. A. Kamakura. *Market Segmentation Conceptual and Methodological Foundations*. Springer Science and Business Media, New York, New York, USA, 2nd edition, 2000.

[203] B. Wellman and M. Gulia. Net-Surfers Don't Ride Alone: Virtual Communities as Communities. In *Networks in the Global Village: Life in Contemporary Communities*, chapter 10, pages 331–366. Westview Press, Boulder, Colorado, USA, 1999.

[204] H. C. White, S. A. Boorman, and R. L. Breiger. Social Structure from Multiple Networks . Blockmodels of Roles and Positions. *American Journal of Sociology*, 81(4):730–780, 1976.

[205] M. Wichmand and S. S. Jensen. Small is beautiful - lurkers engaging through micro-contributions. In *4th European Communication Conference*, 2012.

[206] D. Wiley. Likes, shares, comments: don't be misled by past engagement rates, 1 2021. URL `https://www.thedrum.com/opinion/2021/01/15/likes-shares-comments-don-t-be-misled-past-engagement-rates`.

[207] J. Williams, R. Heiser, and S. J. Chinn. Social media posters and lurkers: The impact on team identification and game attendance in minor league baseball. *Journal of Direct, Data and Digital Marketing Practice*, 13(4), 4 2012. doi: 10.1057/dddmp.2011.44.

[208] J. Woo, J. Son, and H. Chen. An SIR Model for Violent Topic Diffusion in Social Media. In *Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics*, pages 15–19, 2011. doi: 10.1109/ISI.2011.5984043.

[209] Y. Xie, X. Li, E. W. T. Ngai, and W. Ying. Customer Churn Prediction using Improved Balanced Random Forests. *Expert Systems with Applications*, 36(3):5445–5449, 2009. doi: 10.1016/j.eswa.2008.06.121.

[210] C.-L. Yang, C. W. T. Yuan, and H.-C. Wang. When Knowledge Network is Social Network. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–23, 2019. doi: 10.1145/3359266.

[211] S. Yang, F. B. Keller, and L. Zheng. *Social Network Analysis: Methods and Examples*. Sage Publications., 2016.

[212] X. Yang, G. Li, and S. S. Huang. Perceived online community support, member relations, and commitment: Differences between posters and lurkers. *Information & Management*, 54(2), 3 2017. doi: 10.1016/j.im.2016.05.003.

[213] R. K. Yin. *Qualitative Research from Start to Finish*. The Guilford Press, New York, New York, USA, 2nd editio edition, 2016.

[214] D. Zeng, H. Chen, R. Lusch, and S.-H. Li. Social Media Analytics and Intelligence. *IEEE Intelligent Systems*, 25(6):13–16, 2010. doi: 10.1109/MIS.2010.151.

[215] H. Zhang, M. De Choudhury, and J. Grudin. Creepy but Inevitable? The Evolution of Social Networking. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 368–378, 2014. doi: 10.1145/2531602.2531685.

[216] L. Zhao, H. Cui, X. Qiu, X. Wang, and J. Wang. SIR Rumor Spreading Model in the New Media Age. *Physica A: Statistical Mechanics and its Applications*, 392(4):995–1003, 2013. doi: 10.1016/j.physa.2012.09.030.

[217] M. Zimmer. "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology*, 12(4):313–325, 12 2010. doi: 10.1007/s10676-010-9227-5.